UNIVERSITY OF EDINBURGH

College of Science and Engineering

School of Chemistry

Peptide nucleic acid-encoded libraries for microarray-based high-throughput screening

By

Songsak Planonth

Doctor of Philosophy

March 2011

UNIVERSITY OF EDINBURGH

ABSTRACT

College of Science and Engineering
School of Chemistry

Doctor of Philosophy

Peptide nucleic acid-encoded libraries for microarray-based high-throughput screening

By Songsak Planonth

Peptide nucleic acids (PNAs) were used as encoding tags to enable the analysis of peptide libraries by PNA/DNA hybridisation onto DNA microarrays. This allowed entire peptide libraries to be organised and sorted in a two dimensional format whereby all library members could be interrogated and analysed on a one-by-one basis. In this thesis, PNA-encoded peptide libraries, generated by split-and-mix library synthesis, were screened for a variety of functions.

Peptide sequences identified from the screening of a PNA-encoded library were analysed in detail as the first specific substrates for chymopapain. A new PNA-encoded library consisting of D-amino acids was synthesised and screened with a number of proteases in attempts to identify novel/unusual substrates.

PNA-encoded libraries were also used in the screening of peptide libraries for other activities. Thus substrates for catalyst-free Hüisgen cycloaddition were identified following the reaction between an alkyne modified peptide library and azidofluorescein, while cell-penetrating peptides were identified by hybridization of an internalized encoded library onto a DNA microarray.

Declaration of authorship

I, Songsak Planonth, declare that the thesis entitled Peptide nucleic acid-encoded libraries for microarray-based high-throughput screening and the work presented in it are my own.

I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:

  "From 10,000 to 1: Selective synthesis and enzymatic evaluation of fluorescence resonance energy transfer peptides as specific substrates for chymopapain", Diaz-Mochon, J. J.; Planonth, S.; Bradley, M., *Analytical Biochemistry,* **2009,** 384, 101-105.

Signed: _____

Date: _____

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Mark Bradley, for all his guidance and support throughout my PhD, and for giving me this great opportunity of carrying out a PhD in his research group. I also thank the Thai government and Prof. Mark Bradley for financial support during my study.

I would like to thank Dr. JuanJo who provided me with valuable advice and help throughout my PhD. Many thanks go to Géraldine and Dr. Mazen for their great help while I was writing up my thesis in Thailand.

I would like to thank all members of the Bradley group (past and present) who made this time so special and enjoyable. In particular I would like to thank Dr. JuanJo, Dr. Rosario, Luciano, Dr. Delphine and Dr. Adam for the great fun outside the lab. I particularly thank Rahimi and Dr. JuanMa who are always helpful and supportive.

Special thanks to my lab mates, Dr. Alessandra and Frank B in particular, for their valuable advice and giving me a great time working in the lab. Many thanks to Dr. JuanJo, Dr. Delphine, Dr. Effie and Dr. Nina for sharing experiences on the PNA synthesis.

Last but not least, I would like to thank my parents and sisters for their continual support all along my PhD, especially during the hard times.

TABLE OF CONTENTS

ABBREVIATIONS

Amino acids

Three-letter code (one-letter code)

Ala (A)	alanine
Arg (R)	arginine
Asn (N)	asparagine
Asp (D)	aspartic acid
Cys (C)	cysteine
Gln (Q)	glutamine
Glu (E)	glutamic acid
Gly (G)	glycine
His (H)	histidine
Ile (I)	isoleucine
Leu (L)	leucine
Lys (K)	lysine
Met (M)	methionine
Phe (F)	phenylalanine
Pro (P)	proline
Ser (S)	serine
Thr (T)	threonine
Trp (W)	tryptophan
Tyr (Y)	tyrosine
Val (V)	valine

General

AA	amino acid
Abl	Abelson tyrosine kinase
Abu	amino butyric acid
Ac	acetyl
Ahx	6-aminohexanoic acid
Alloc	allyloxycarbonyl
Ar	aryl (NMR)
ATP	adenosine triphosphate
Bhoc	benzhydryloxycarbonyl
BLAST	basic local alignment search tool
Boc	*tert*-butoxycarbonyl
br	broad (NMR)
Cbz	carboxybenzyl
Cp*	1,2,3,4,5-Pentamethylcyclopentadiene
CPP	Cell-penetrating peptide
d	doublet (NMR)
DAP	diaminopimelate
DCC	dicyclohexylcarbodiimide
DCM	dichloromethane
dd	double doublet (NMR)
Dde	*N*-[1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl]
DIC	diisopropylcarbodiimide
DIPEA	*N, N*-diisopropylethylamine

DMAP	4-dimethylaminopyridine
DMF	*N,N*-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetatic acid
ELSD	evaporative light scattering detector
ESI	electrospray ionisation
FAM	fluorescein amide
FISH	fluorescence in situ hybridisation
Fmoc	9-fluorenylmethoxycarbonyl
FRET	fluorescence resonance energy transfer
GlcNAc	*N*-acetylglucosamine
GPCR	G protein coupled receptor
HOBt	1-hydroxy-1H-benzotriazole
HPLC	high performance liquid chromatography
HTS	high-throughput screening
IgG	immunoglobulin G
IR	infrared
J	scalar coupling constant (NMR)
Llp	lysine-like peptoid monomer
m	multiplet (NMR) or medium (IR)
m/z	mass/charge ratio (mass spectrometry)
MALDI-TOF	matrix-assisted laser desorption/ionisation time-of-flight
Mmt	monomethoxytrityl

mRNA	messenger ribonucleic acid
MS	mass spectrometry
MurNAc	*N*-acetylmuramic acid
NEM	N-ethyl morpholine
NMP	*N*-methyl-2-pyrrolidone
NMR	nuclear magnetic resonance
PBS	phosphate buffered saline
PEG	polyethyleneglycol
PEGA	polyethylene glycol dimethyl acrylamide
PNA	peptide nucleic acid
Pur	purine
PyBOP	benzotriazol-1-yloxytri(pyrrolidino)-phosphonium hexafluorophosphate
q	quartet (NMR)
RFU	relative fluorescence units
RHOD	rhodamine B amide
RNA	ribonucleic acid
s	singlet (NMR) or strong (IR)
SDS	sodium dodecyl sulfate
SPPS	solid phase peptide synthesis
SSC	sodium chloride/sodium citrate
StdDev	standard deviation
t	triplet (NMR)
TAMRA	tetramethyl-6-carboxyrhodamine

td	triplet of doublets (NMR)
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIS	tri*iso*propylsilane
TLC	thin layer chromatography
T_m	melting temperature
tR	retention time
UV	ultraviolet
w weak	(IR)
δ	chemical shift (NMR)

Chapter 1 Introduction.

This thesis presents the application of peptide nucleic acids (PNAs) in high-throughput screening. It starts with an introduction to PNAs, the concept of split-and-mix synthesis and a review of current encoding methodologies. Finally, the synthesis of PNA-encoded libraries that will be used throughout this thesis is also presented.

In chapter 3, an example of a PNA-encoded high-throughput peptide library screening is demonstrated. The first specific substrates for chymopapain, which were identified by this method, were studied and their kinetic parameters were determined. In addition, other two encoded libraries were screened as potential substrates for chymopapain exploring D-amino acid-containing peptides. It becomes interesting to screen for D-amino acid-containing peptides as specific substrates for proteases due to the possible applications that such novel sequences would have, and a high-throughput screening method would be the most powerful approach of being able to detect/find them. Two encoded D-amino acid containing peptide libraries were therefore screened to find substrates of two proteases, chymopapain and a newly discovered protease – gumby.

In chapter 4, screening of the PNA-encoded peptide library as potential substrates for self-promoting Hüisgen cycloaddition is presented. The characteristics of the reaction are discussed. While Cu(I) is the catalyst for the reaction, the importance of the absence of metal catalyst, especially in biological system, is essential. After library screening under the catalyst-free conditions and data analysis, initial peptides of interest are identified.

Chapter 5 thesis introduces cell penetrating peptides (CPPs) and the screening of PNA-encoded library for novel CPPs. It starts with the delivery of known peptides capable of cellular delivery conjugated to PNA oligomers. The PNA-encoded peptide library is tested with two different cell lines (HeLa and B16F10) and the internalized sequences are analysed in a high-throughput manner. The “hit” peptides as cellular delivery agents are synthesised and their cellular delivery is evaluated and compared with known delivery agents.

Chapter 2: Synthesis of PNA monomers and PNA-encoded peptide libraries.

2.1 Introduction to peptide nucleic acids.

Peptide nucleic acids (PNAs) were first reported in 1991 by Nielsen as a DNA mimic in which the sugar-phosphate backbone of DNA had been replaced by a repeating polyamide chain of *N*-(2-aminoethyl)glycines (**Figure 2.1**)[1]. PNAs can hybridise efficiently to complementary DNA or RNA sequences according to standard base paring rules. Due to the lack of charge on the PNA backbone, binding affinities of PNA toward DNA and RNA, under physiological conditions, are higher than those of DNA to DNA, due to the presence of two negatively charged phosphate diester backbones in the latter instance. PNAs are resistant to degradation by nucleases and proteases and because of their pseudo peptidic backbone, their physical properties are more akin to peptides than nucleic acids.

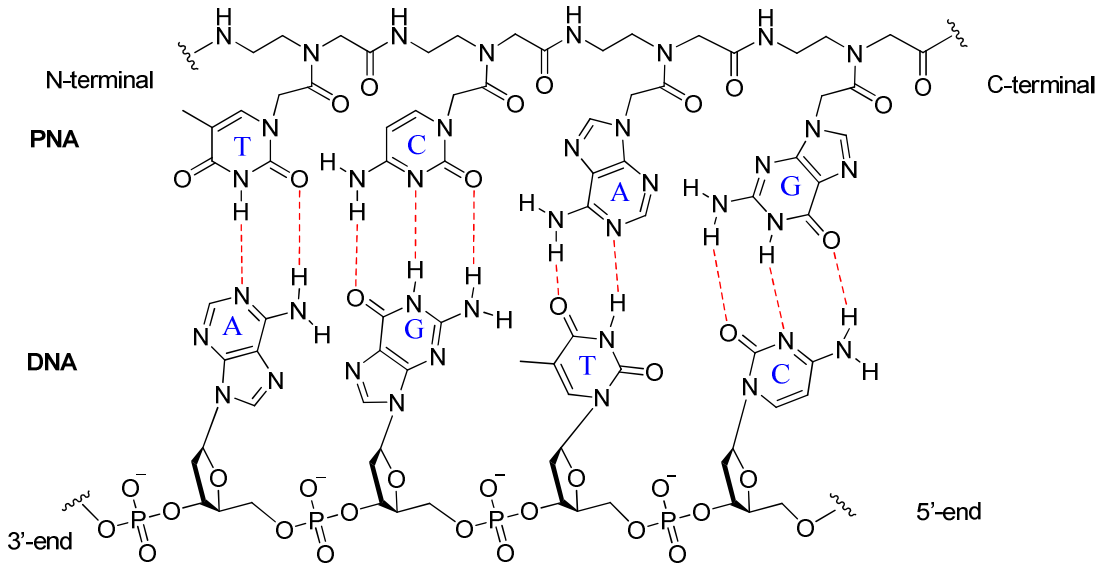


Figure 2.1: Hybridisation between PNA and DNA.

PNA (due to its chemical make up) can hybridise to DNA in either orientation, however the prefered orientation in the hybridisation of PNA to DNA is with amino terminus of the PNA facing the 3'-end of DNA, as shown in **Figure 2.1** (the so-called antiparallel orientation). The alternative orientation results in about a 13°C difference in melting temperature (T_m) for a PNA/DNA duplex and 21°C for a PNA/RNA duplex[2].

2.2 Applications of peptide nucleic acids.

2.2.1 Antisense therapy.

Antisense therapy is based on the treatment of genetic disorders by preventing translation of mRNAs into proteins. Antisense-oligonucletides (AS-NOs, modified oligodeoxynucleotides) often express their antisense activity in two major mechanisms. On the one hand, AS-NOs are designed to activate RNase, enzyme which cleave RNA moiety of a DNA-RNA duplex, on the other hand, they inhibit translation of the mRNA at ribosome by steric blockade [3]. PNAs are one type of the AS-NOs. Since PNA-RNA hybrids are not substrates of enzyme and they have high binding efficiencies, their mode of action is believed to be translation-blocking rather than inducing RNA degradation[4,5].

2.2.2 Fluorescent *in situ* hybridisation.

Fluorescent *in situ* hybridisation (FISH) is a method by which fluorescently labeled probes target specific sequences in a cell. Using DNA probes, a number of problems can be faced, for example, degradation of probes by nucleases and non-specific binding.

PNA is one of the DNA mimics that have been used in FISH. PNA/DNA duplexes have higher thermal stability compared to DNA/DNA duplexes and therefore shorter sequences are required. In addition, the destabilising effect of a single-base mismatch in PNA/DNA duplexes is higher than that in DNA/DNA duplexes, resulting in higher specificities of detection[6]. Examples of this technique are the detection and identification of bacteria existing in the biofilms of chronic wounds[7]. Indeed a commercial system is now available based on PNA hybridisation for bacterial detection.

2.2.3 Nucleic acid capture.

PNAs have been used as tools in nucleic acid purification and recovery (which is often performed using DNAs). PNA probes show much more rapid and efficient binding/recovery of the target nucleic acids compared to DNA probes under low salt conditions. These probes have also been found to capture target DNAs at concentrations as low as zeptomolar ($x10^{-21}$ M)[8], while PNA-loaded beads have been used to capture RNA in an automated manner with higher efficiency than DNA beads[9].

2.3 Synthesis of peptide nucleic acids building blocks.

2.3.1 Protecting groups used in peptide nucleic acid monomers synthesis.

The synthesis of PNA oligomers is performed by the coupling of building blocks, so called monomers, based on derivatives of *N*-(2-aminoethyl)glycine. Four natural nucleobases are used with three of these building blocks containing two amino groups (an exocyclic amine and the *N*-(2-aminoethyl)glycine unit) (**Figure**

2.2). Orthogonal protecting groups are therefore (like for DNA synthesis) necessary for PNA building block and many combinations have been used to date.

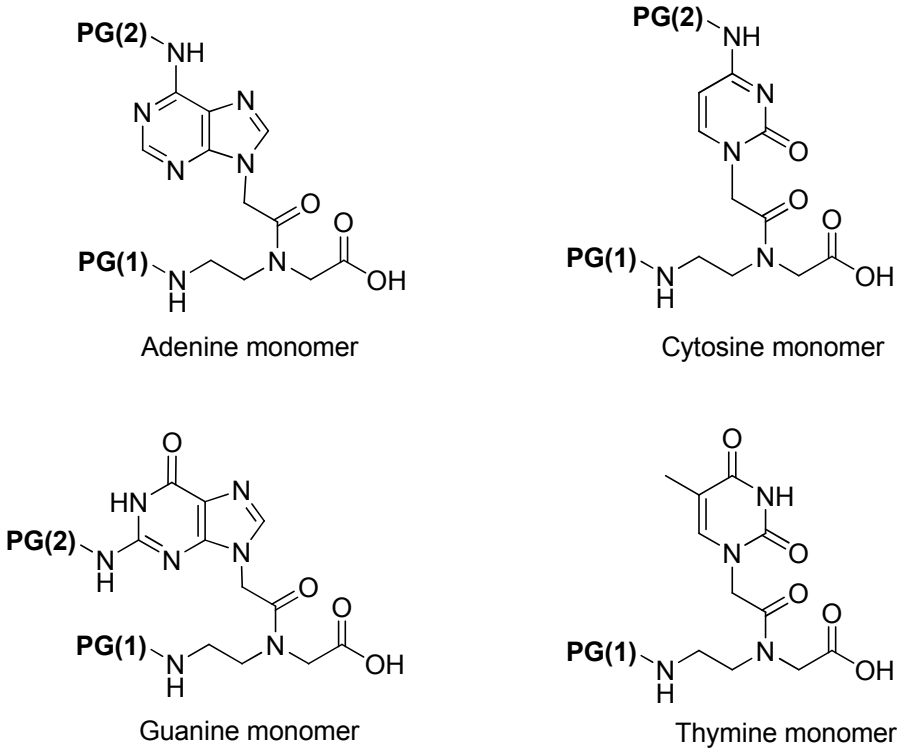


Figure 2.2: PNA monomers of four nucleobases.

The combination of acid labile Boc and the less acid labile carboxybenzyl (Cbz) were introduced in 1994[10]. Boc being removed by treatment of 50% TFA while the Cbz group could be removed by treatment with hydrogen fluoride (see **Figure 2.3**).

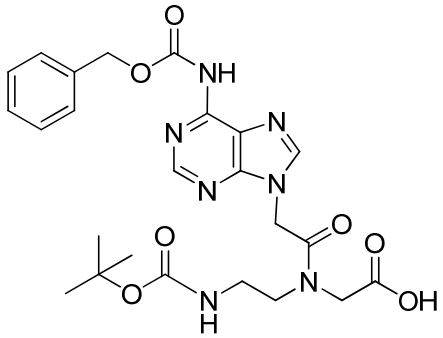


Figure 2.3: Adenine monomer protected with Boc (chain terminal) and Cbz (side chain).

Basic-labile 9-fluoren-9-ylmethoxycarbonyl (Fmoc) and Cbz was a combination introduced in 1995 (**Figure 2.4**)[11]. In addition to preventing undesired side reactions in the oligomerization, the Cbz protecting group has also found to increase monomer solubility.

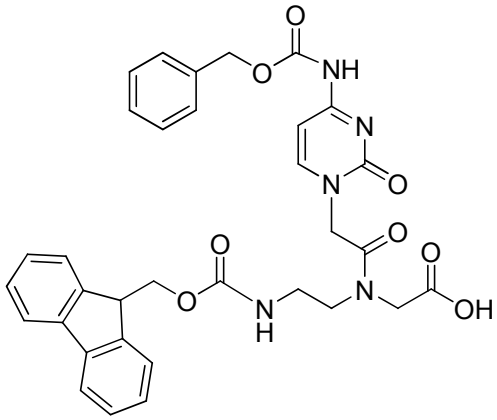


Figure 2.4: Cytosine monomer protected with Fmoc and Cbz.

The combination of Fmoc and the acid-labile mono-methoxytrityl (Mmt) group was introduced in 1996 (**Figure 2.5**)[12]. PNA monomers with these protecting groups can be used in automated peptide synthesizers with standard Fmoc deprotection chemistry to give oligomers. Final treatment with 95% TFA giving Mmt deprotection, and cleavage from the solid support.

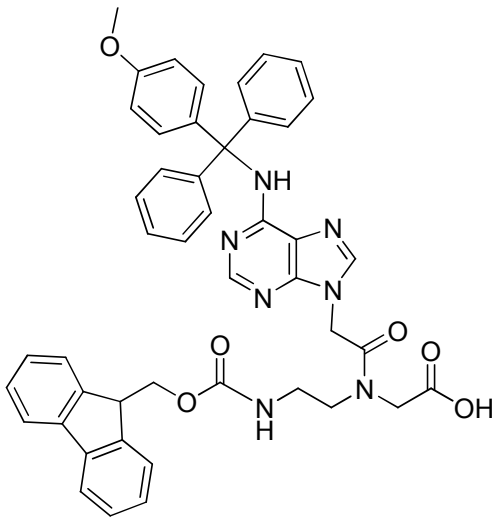


Figure 2.5: Adenine monomer protected with Fmoc and Mmt.

Basic-labile Fmoc and acid-labile benzhydroloxycarbonyl (Bhoc) protected PNA monomers have been successfully used in solid phase synthesis (**Figure 2.6**). The Bhoc group (as for the Mmt) being removed by treatment with mild (relatively) acid[13].

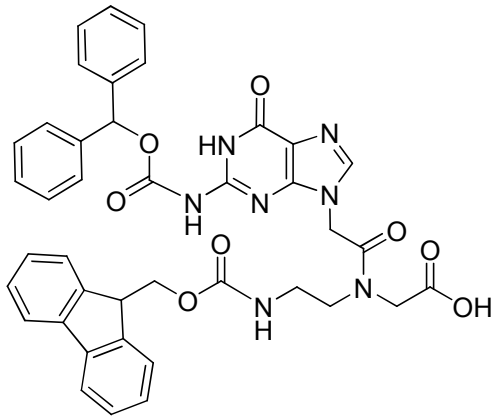


Figure 2.6: Guanine monomer protected with Fmoc and Bhoc.

The Bradley group has introduced a technique where PNA could be synthesised using monomers orthogonal to Fmoc-amino acids[14-16]. The PNA monomers designed are shown in **Figure 2.7**[17] which the amino terminus protected with 1-(4,4-dimethyl-2,6 dioxacyclo-hexylidene) ethyl (Dde) and the exocyclic

amino groups of the nucleobases protected (if necessary) with acid-labile mono-methoxytrityl (Mmt) groups.

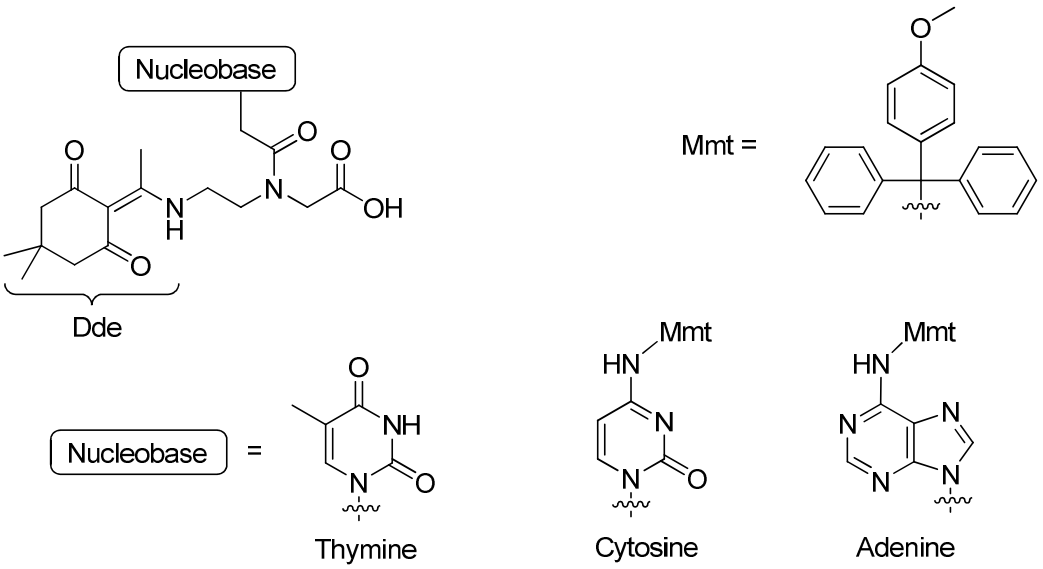
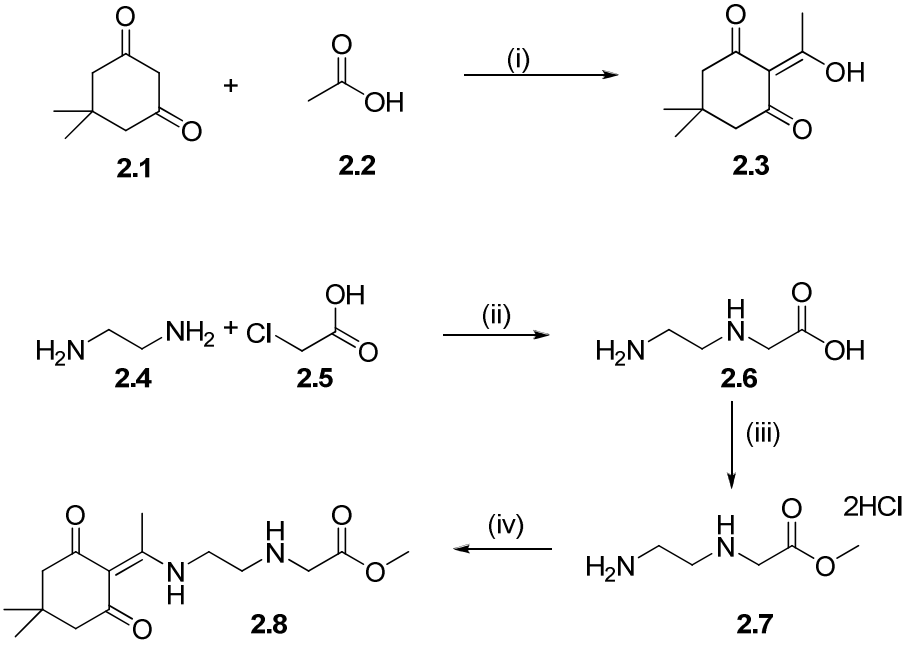


Figure 2.7: PNA monomers for the synthesis of PNA-peptide conjugate.

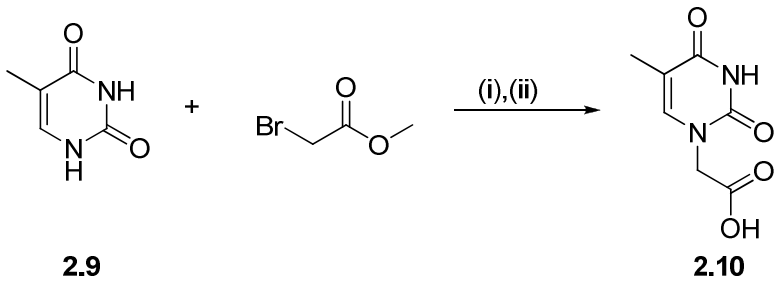
The synthesis of these monomers starts with the synthesis of the Dde protected backbone as shown in **Scheme 2.1**.



Scheme 2.1: Synthesis of Dde-protected backbone; *Reagents and conditions*: (i) 1 equiv. DCC, 0.1 equiv. DMAP, DMF 60°C, 16 h; (ii) 16 h; (iii) 10 equiv. $SOCl_2$, MeOH, reflux, 16 h.; (iv) 2 equiv. DIPEA, 1 equiv. **2.3**; DCM/EtOH (1/1), 16 h.

Thus Dde-OH **2.3** was prepared from 5,5-dimethyl-1,3-cyclohexanedione **2.1** and acetic acid **2.2** in the presence of DMAP (0.1 equiv.) and DCC (1.0 equiv.) with gentle heat. Ethylendiamine **2.4** was alkylated with chloroacetic acid **2.5**, giving 2-aminoethylglycine **2.6** which was esterified to give ester **2.7**. The *N*-terminus was protected by reacting with Dde-OH **2.3**, giving methyl *N*-{2-[1-(4,4-dimethyl-2,6-dioxo-cyclohexyliden)-ethylamino]-ethyl}-glycinate **2.8**.

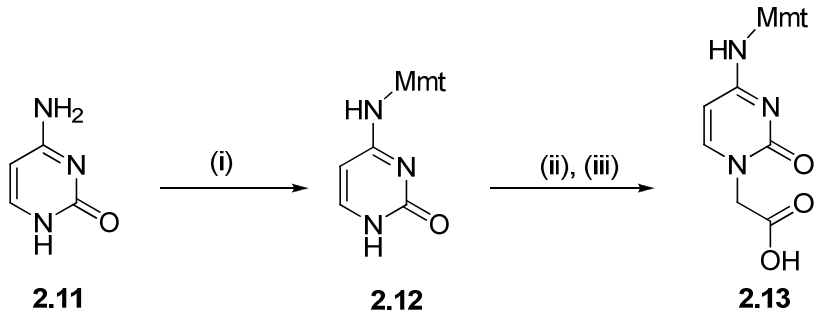
The nucleobases were alkylated with methyl bromoacetate and, for those with exocyclic amino groups, protected with the Mmt group. Thus (**Scheme 2.2**) thymine **2.9** was alkylated in DMF with methyl bromoacetate using K_2CO_3, and the resulting methyl ester was then saponified with NaOH to give thymin-1-yl acetic acid **2.10**.



2.9 **2.10**

Scheme 2.2: Thymine derivatization; *Reagents and conditions*: (i) 1 equiv. K_2CO_3, DMF, N_2 atm, 16 h; (ii) 10% NaOH (aq), reflux, 10 min.

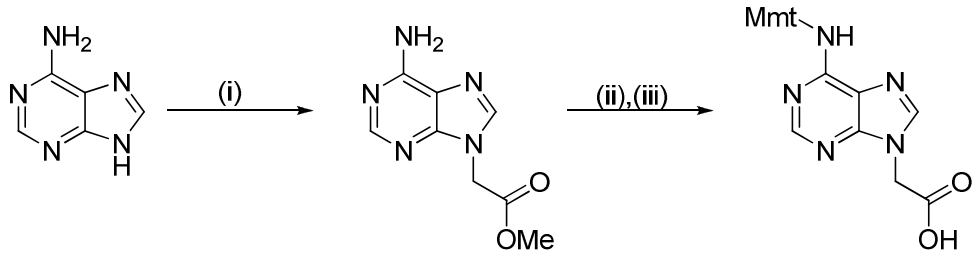
Cytosine **2.11** was initially protected by reacting with Mmt-Cl in pyridine. Mmt-protected cytosine **2.12** was then alkylated similarly to thymine to give [N^4-(4-methoxytrityl)-cytosin-1-yl]-acetic acid **2.13** (**Scheme 2.3**).



2.11 **2.12** **2.13**

Scheme 2.3: Cytosine derivatization; *Reagents and conditions*: (i) 1.5 equiv. Mmt-Cl, 1.0 equiv. NEM, Pyr, 40 ºC, overnight; (ii) 1 equiv. K_2CO_3, 1.0 equiv. methyl 2-bromoacetate, DMF, N_2 atm., overnight; (iii) 10% NaOH (aq), reflux, 10 min.

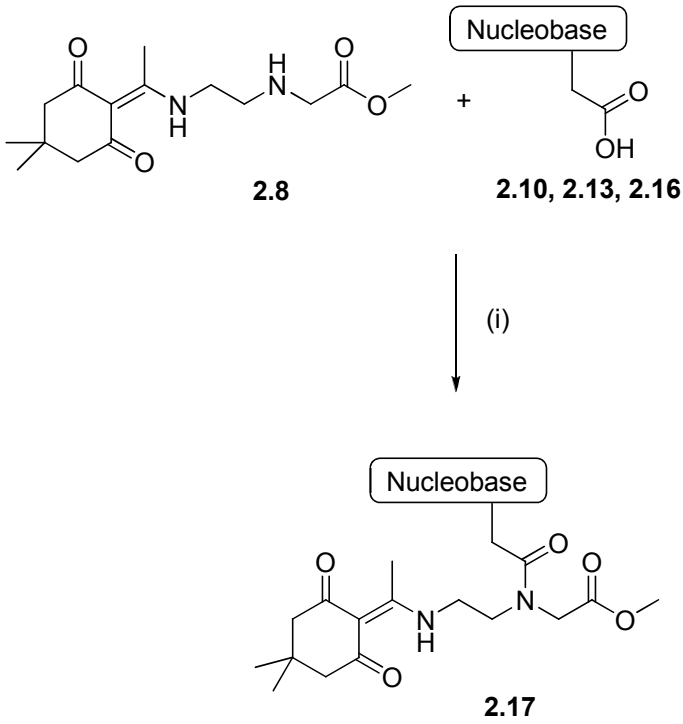
The adenine derivatives were prepared in a slightly different manner to those of cytosine. Adenine **2.14** was firstly alkylated (**2.15**), prior to protection (as for **2.12**) with [N^6-(4-Methoxytrityl)-adenin-9-yl]-acetic acid **2.16** obtained after saponification with NaOH (**Scheme 2.4**).



2.14 **2.15** **2.16**

Scheme 2.4: Adenine derivatization; *Reagents and conditions*: (i) 1.1 equiv. NaH, 1 equiv. methyl 2-bromoacetate, DMF, N_2, 4 h; (ii) 1.5 equiv. Mmt-Cl, 1 equiv. NEM, DCM, 40°C, 3 h, then 25°C, 16 h; (iii) 1 N aqueous NaOH, reflux, 2 h.

Once the Dde-protected backbone and nucleobase derivatives were available, these two parts were coupled using DCC and HOBt (**Scheme 2.5**). Yields for these 3 nucleobase are given in **Table 2.1**.



Scheme 2.5: Coupling of Dde-protected backbone and derivatized nucleobases; *Reagents and conditions*: (i) 1.2 equiv. DCC, 1.0 equiv. HOBt, 16 h.

Saponification of esters **2.17 (A, C, T)** was carried out using 1 M Cs_2CO_3 in 1:1 MeOH/H_2O, with PNA monomers obtained in the range of 73-85% from ester **2.17** (**Table 2.1**).

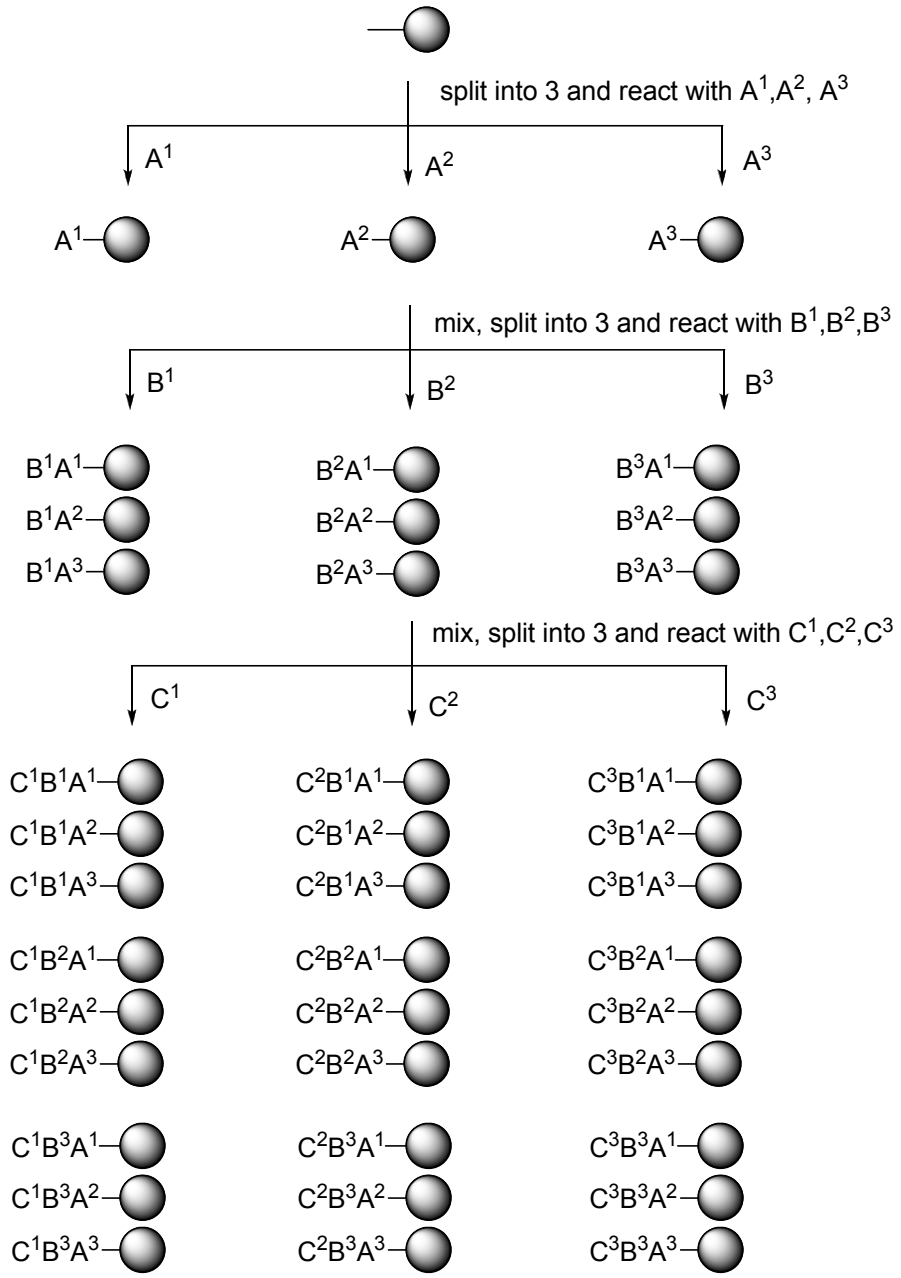
Nucleobase	% Yield (2.17)	% Yield (monomer)
Thymine	92	73
Cytosine(Mmt)	69	80
Adenine(Mmt)	53	85

Table 2.1: Yields of coupling of Dde-protected backbone and different derivatized nucleobases (**2.10**, **2.13**, **2.16** to **2.17**) and PNA monomers.

2.4 Synthesis of peptide nucleic acids-encoded peptide libraries.

2.4.1 Split-and-mix synthesis.

Split-and-mix synthesis was reported by Furka as a method to synthesise libraries of peptide mixtures using solid phase synthesis (**Scheme 2.6**)[18]. This method starts with dividing the solid support into *m* portions and they are reacted with *m* different building blocks (amino acids). After reaction, the portions are mixed into one pool and divided again into *m* portions (each portion contains *m* different compounds). These react with a second building block, and the process is repeated until the required length of the peptide is reached. While the reaction number increases linearly ($m \rightarrow 2m \rightarrow 3m \rightarrow nm$), the number of products increases exponentially ($m \rightarrow m^2 \rightarrow m^3 \rightarrow m^n$).



Scheme 2.6: Split-and-mix synthesis. The solid support is devided into 3 portions (m = 3) and reacted for 3 rounds (n = 3), giving 27 (3^3) products from 9 (3x3) reactions.

Although this approach can produce potentially millions of peptides (each attached to a bead), the identity of the peptide on a bead can be difficult to determine. This therefore led to methods of bead encoding.

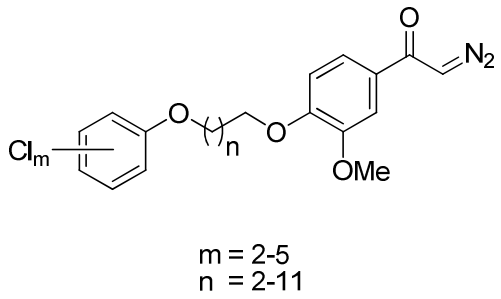
2.4.2 Encoding methods in combinatorial chemistry.

2.4.2.1 DNA encoding.

In 1992, Needels[19] and Brenner[20] introduced a method to identify active peptides from a combinatorial peptide libraries using "genetic" tags. A unique DNA tag for each amino acid was chemically attached to solid support in the split-and-mix library synthesis, resulting in each peptide being encoded by a unique DNA sequence. The active peptides/beads had their DNA sequences amplified by PCR with nucleic acid sequencing revealing the peptide sequences. In one approach[19] peptide sequences on the solid support (with the DNA encoded peptide library) were screened for binding with a fluorescently labeled antibody. After analysis using a fluorescence-activated cell sorting instrument, the oligomers attached to the fluorescent beads were amplified and analysed, revealing the peptide sequences on the beads.

2.4.2.2 Halophenol encoding.

Still developed halophenol tags (**Figure 2.8**) which were detected by electron capture gas chromatography[21]. Tagging reagents were attached to the solid support by a diazoketone functional group being converted to a reactive acylcarbene by rhodium trifluoroacetate. Every building block in each step was encoded by a unique combination of tags and the peptides/small molecules on the beads were identified by the analysis of the tags released from the beads upon oxidation.



m = 2-5
n = 2-11

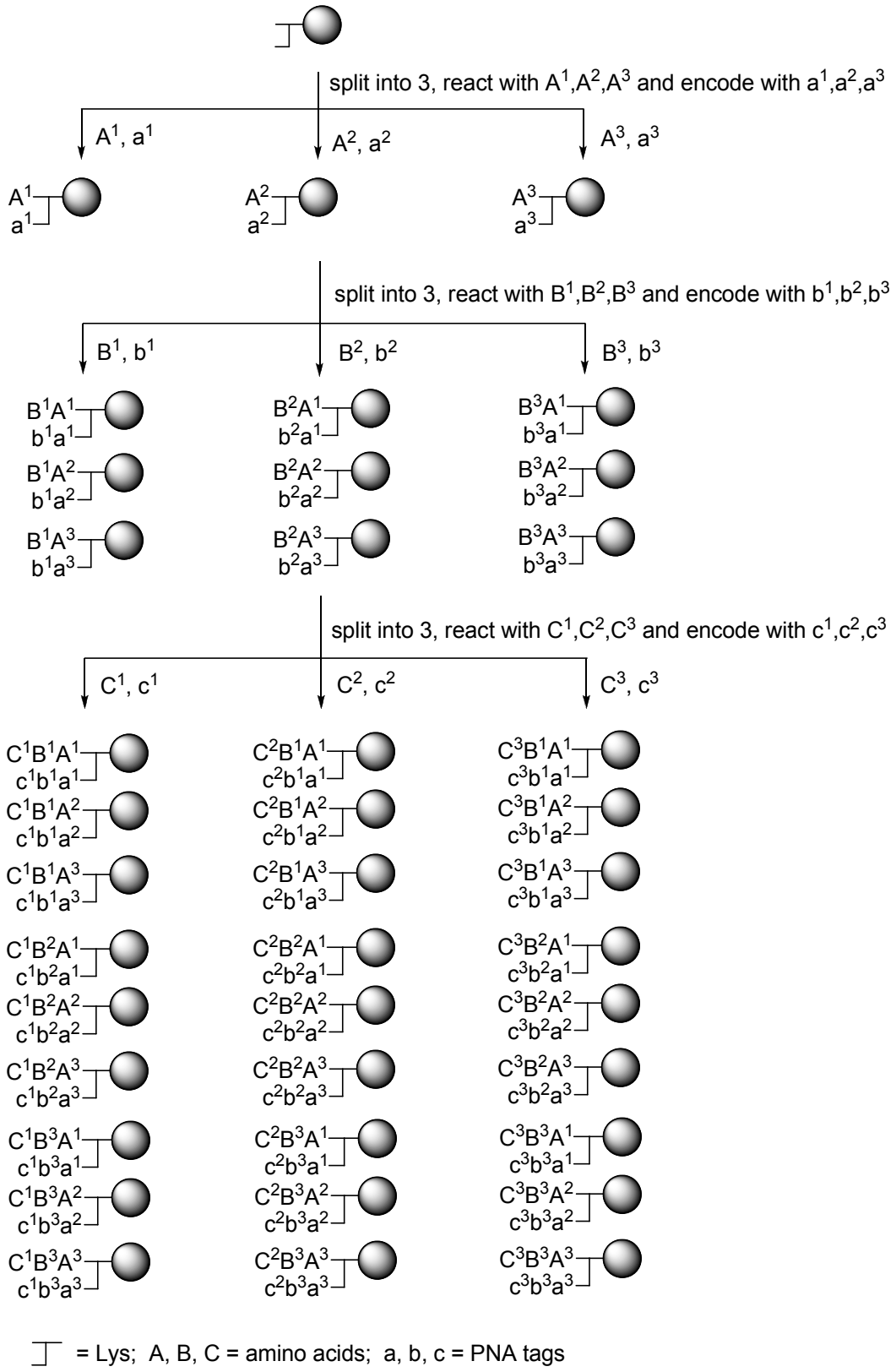
Figure 2.8: General structure of halophenol tags.

2.4.2.3 Colloidal barcodes encoding.

Colloidal barcoding technique is a method to encode the split-and-mix library introduced by Trau[22]. In this method, fluorescent silica particles (1 μm in diameter) were physically immobilised onto the solid support. Each building block in every round was encoded by the unique combination of fluorescent silica particles and the libraries were decoded using fluorescent microscopy

2.4.2.4 PNA-encoded combinatorial libraries and DNA microarrays.

The Bradley[23], Schultz [24] and Winssinger[25] research groups developed the strategy of using PNAs as tags in split-and-mix library synthesis. Using PNA/DNA hybridisation properties, the PNA-encoded libraries could be analysed in a high-throughput manner using DNA microarrays with all library members analysed in one screen. This encoding method, in which the tag is directly attached to the peptide, was used throughout this thesis and it is summarized in **Scheme 2.7**, in which PNA codes were conjugated to the peptides *via* a lysine residue and the library attached to the solid support using an acid-cleavable Rink amide linker (H-Rink-PEGA, **Figure 2.9**).



Scheme 2.7: PNA encoded split-and-mix synthesis.

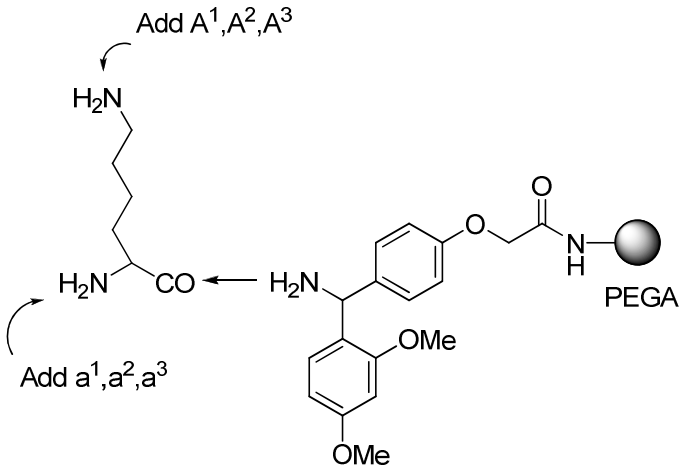
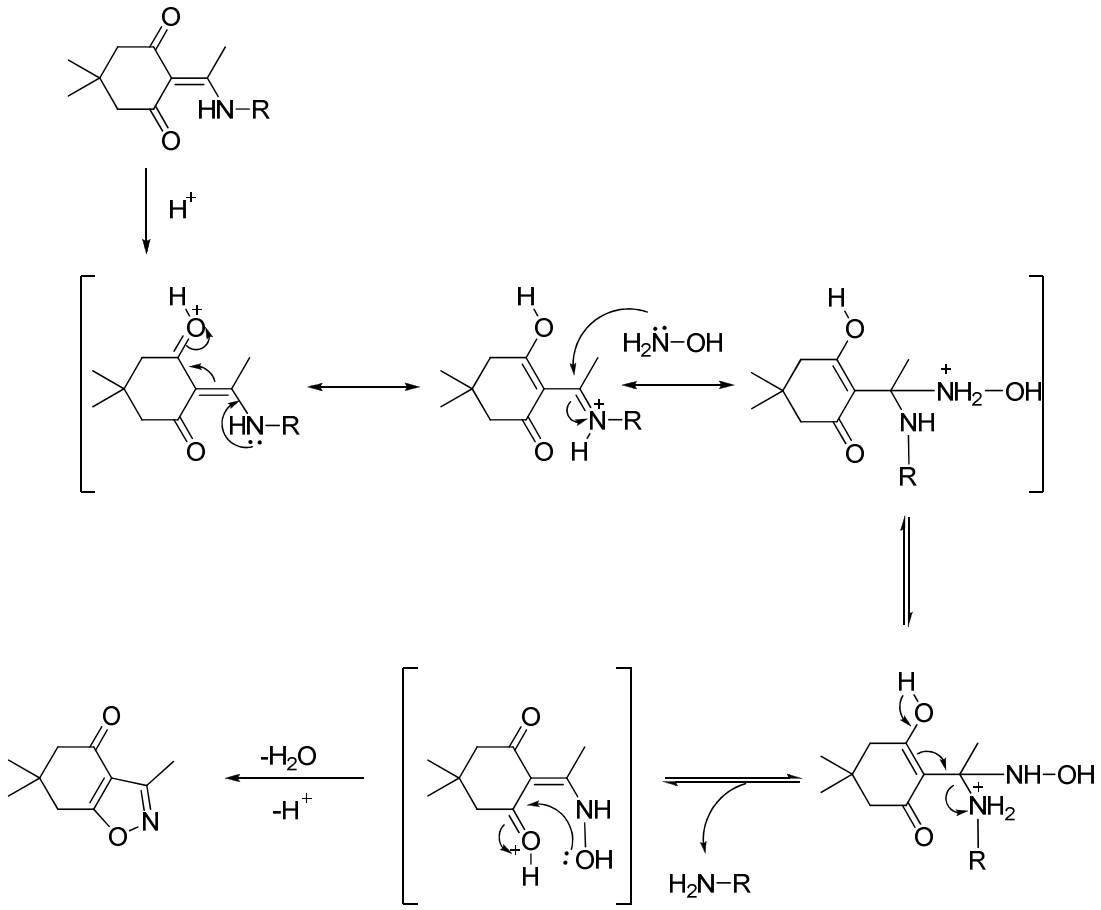
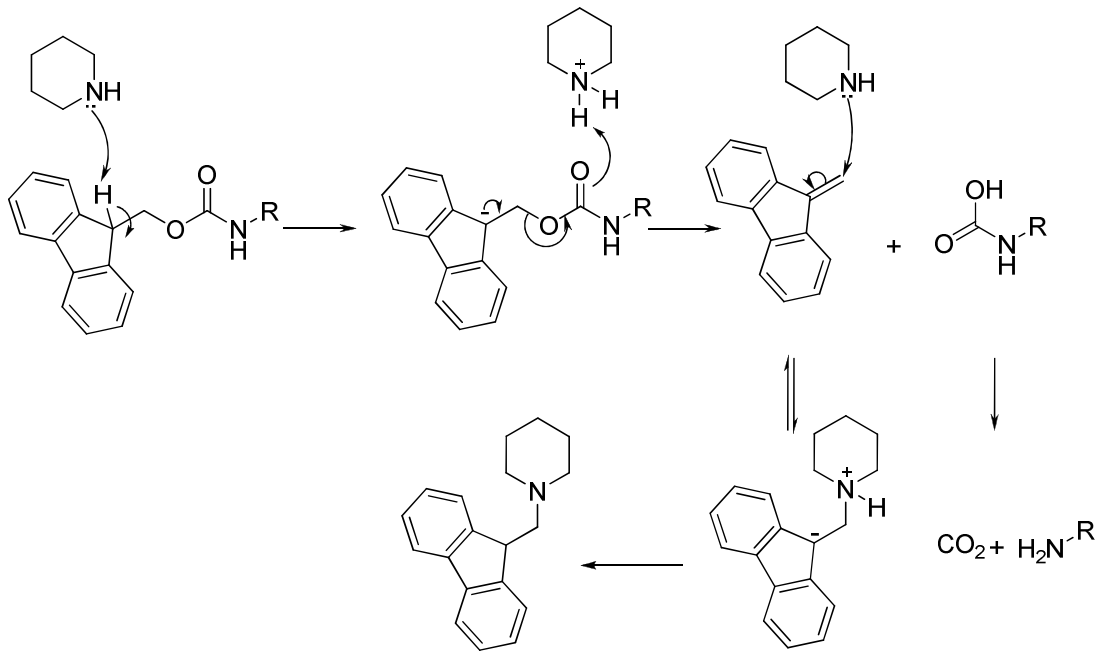


Figure 2.9: Structure of Rink amide linker installed on PEGA resin, with the tri-fucntional nature of Lys used for library synthesis, encoding and solid-phase attachment.

For encoded split-and-mix synthesis, it was necessary that the introduction of tags (a, b, c in **Scheme 2.7**) must not interfere with the introduction of the library building blocks (A, B, C in **Scheme 2.7**) and *vice versa* therefore protecting groups on the tags and library building blocks must be orthogonal. The Bradley group has developed a robust protecting group cleavage strategy that is orthogonal to Fmoc and Boc, using the Dde protecting group. The Dde group can be deprotected by hydroxylamine under slightly acidic conditions (mixture of $NH_2OH{\cdot}HCl$ and imidazole) (**Scheme 2.8**) whilst the Fmoc group is deprotected under basic conditions (**Scheme 2.9**).



Scheme 2.8: Dde deprotection using $NH_2OH{\cdot}HCl$/imidazole.



Scheme 2.9: Fmoc deprotection using piperidine.

In this thesis a 625 member library was encoded with a 3 letter code, with three PNA monomers, A, C and T[14]. For the 1296 member libraries, a 4 letter code composed of A, C and T monomers was used. This code, developed by Brenner[26], allows all the members to be isothermal and has the advantages that only three PNA monomers: A, C and T, were required for library synthesis.

2.4.3 Synthesis of a 625 member FRET-based D-amino acid-containing peptide library.

The general structure of the FRET based D-amino acid containing library is shown in **Figure 2.10**, in which L-Lys was used for the conjugation of the PNA tags and D-Lys for the conjugation of fluorescein.

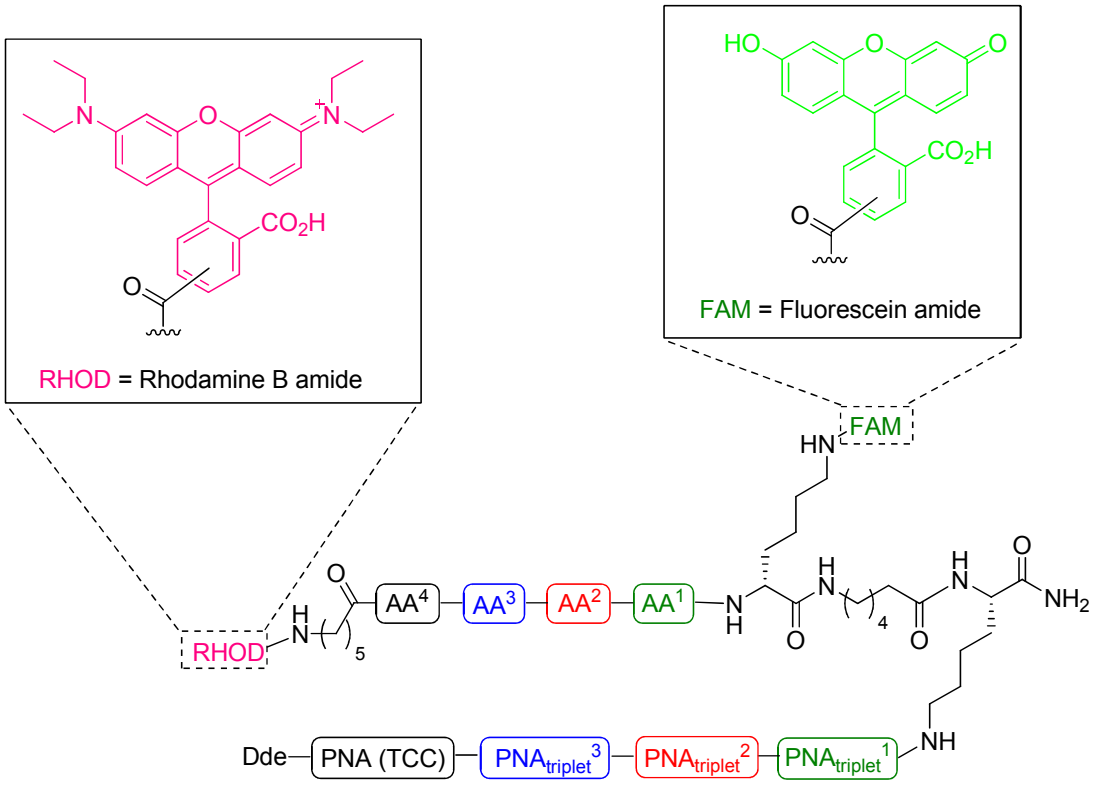
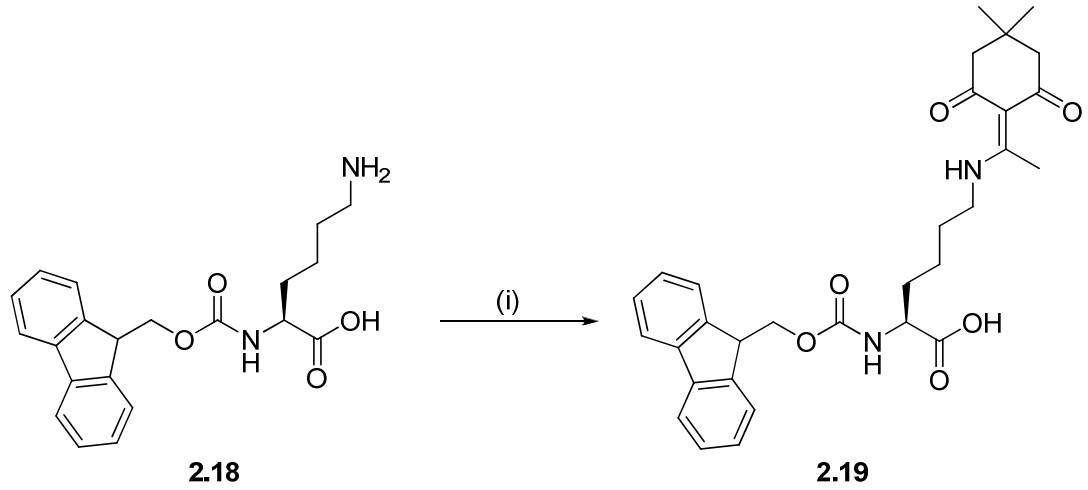


Figure 2.10: Library **L 2.1**.

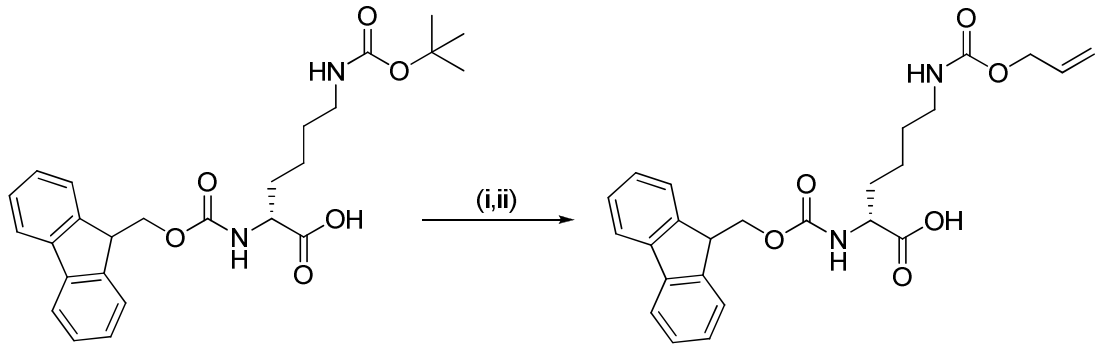
Fmoc-L-Lys(Dde)-OH (**2.19**) was synthesised from Fmoc-L-Lys(H)-OH (**2.18**) and Dde-OH (**2.3**) with a catalytic amount of TFA in refluxing ethanol (**Scheme 2.10**)[27].



2.18 **2.19**

Scheme 2.10: Synthesis of Fmoc-L-Lys(Dde)-OH (**2.18**); *Reagents and conditions*: (i) 1.0 equiv Dde-OH (**2.3**), 0.1 equiv. TFA, EtOH, reflux[27].

Fmoc-D-Lys(Alloc)-OH (**2.21**) was synthesised from Fmoc-D-Lys(Boc)-OH (**2.20**). The butyloxycarbonyl group was removed by the treatment with TFA followed by the treatment with commercially available allyl 1-benzotriazolyl carbonate in the presence of TEA (**Scheme 2.11**).



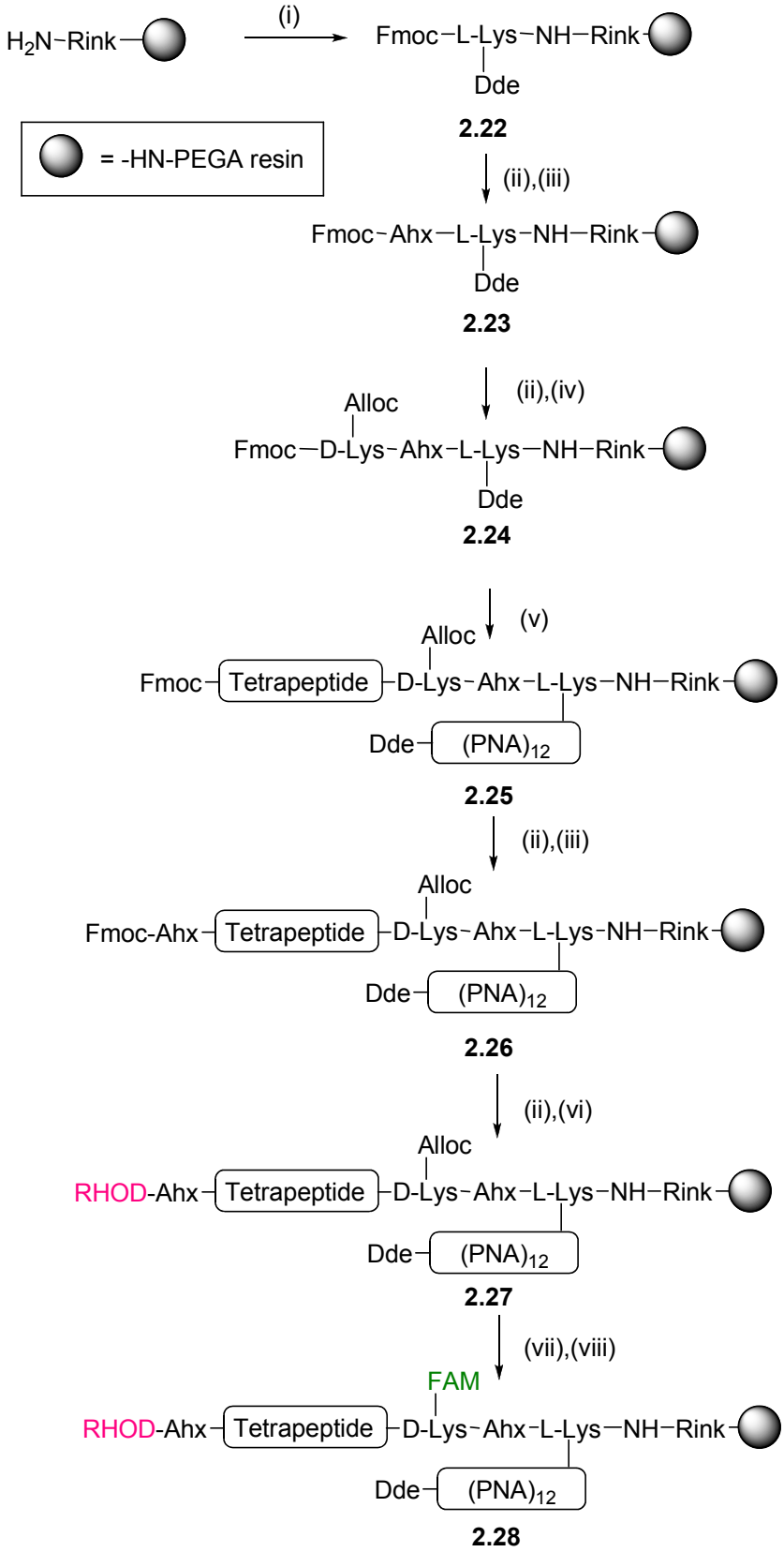
2.20 **2.21**

Scheme 2.11: Synthesis of Fmoc-D-Lys(Alloc)-OH (**2.21**); *Reagents and conditions*: (i) 1:1 TFA:DCM, 2 h; (ii) 1.0 equiv. allyl 1-benzotriazolyl carbonate, 2.0 equiv. TEA, 2.5 h.

The library synthesis scheme is summarized in **Scheme 2.12**. All coupling steps were performed under microwave irradiation, using benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) as a coupling reagent. H-Rink-PEGA resin (**Figure 2.9**) was first coupled to Fmoc-L-Lys(Dde)-OH (**2.19**) followed by Fmoc protected 6-aminohexanoic acid (Fmoc-Ahx-OH) as a spacer (**2.23**). After the removal of the Fmoc group, Fmoc-D-Lys(Alloc)-OH (**2.21**) was coupled and the resulting resin **2.24** was then ready for split-and-mix synthesis.

The encoded split-and-mix synthesis was performed following **Scheme 2.7**. Every position on the peptide was varied with 5 different amino acids. L-Ala was included as one of building blocks since it was found to be the most preferable amino acid cleaved by chymopapain (see section **3.1** for detail). For the first three amino acids (AA^1 to AA^3), every amino acid was encoded with unique PNA triplet according to **Table 2.2**. For amino acids at position 4 (AA^4), all amino acids presented in **Table 2.2** were used, but every one was encoded with the same PNA triplet code (TCC) and the library was left split in 5 portions. As a result, 125 PNA oligos could be used to encode the whole library of 625 members and therefore only 125 DNA probes would be needed in the analysis.

A chromophoric pair was introduced into the library (**2.25**). After Fmoc deprotection, the Fmoc-Ahx-OH spacer was coupled (**2.26**), followed by 5(6)-carboxyrhodamine B, (**2.27**). The Alloc, protecting group on the N^{ε} of D-Lys, was removed by tetrakis(triphenylphosphine)palladium(0). The resulting resin was coupled with 5(6)-carboxyfluorescein, giving a Rhodamine B amide (RHOD) and Fluorescein amide (FAM) labeled tetrapeptide library (**2.28**).

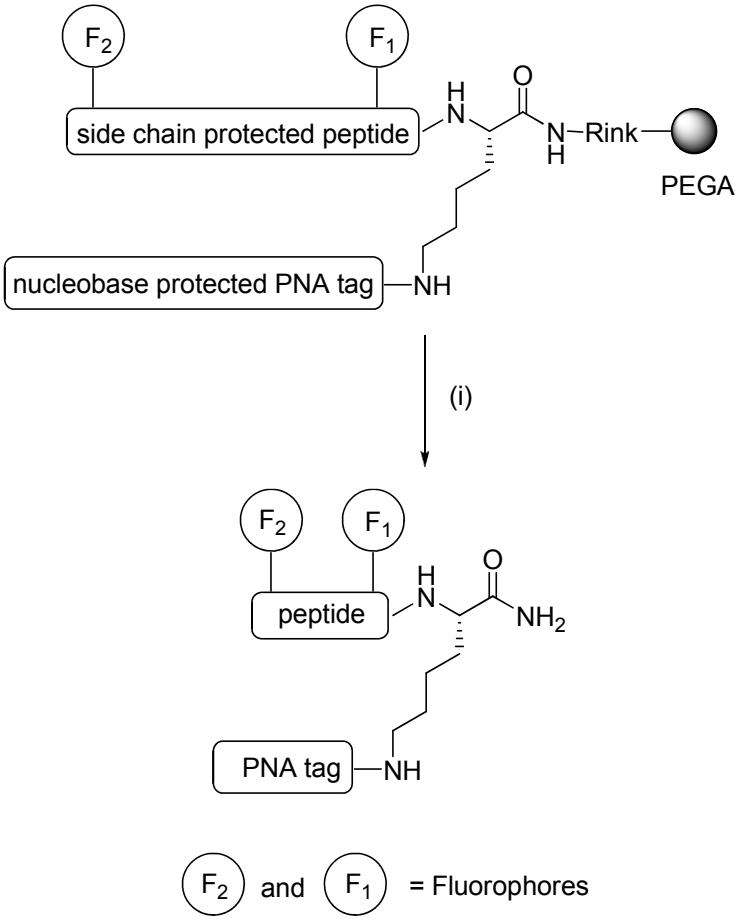


Scheme 2.12: Synthesis of a PNA-encoded tetrapeptide library consisting of D-amino acid (**L 2.1**); *Reagents and conditions:* (i) **2.19**; (ii) 20% piperidine / DMF, 2 x 10 min; (iii) Fmoc-Ahx-OH; (iv) **2.21**; (v) Split-and-mix library synthesis; (vi) 5(6)-carboxyrhodamine B; (vii) 12 equiv. $PhSiH_3$, 0.1 equiv. $Pd(PPh_3)_4$ / dry DCM, 2 x 15 min; (viii) 5(6)-carboxyfluorescein; (5.0 equiv. amino acids, PNA monomers or dye, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, microwave irradiation).

Amino acids	PNA code (*N* to *C*)
Gly	TAA
D-Ala	ATA
L-Ala	TCC
D-Met	CTC
D-Glu	AAC

Table 2.2: Amino acids and their PNA triplet codes used in the split-and-mix library synthesis in positions AA^1 to AA^3.

The cleavage of the PNA encoded library, as well as the removal of the acid labile protecting groups on the amino acid side chain and PNA nuclebases was achieved in one step by the treatment of the library with a mixture of TFA, triisopropylsilane (TIS) and DCM (**Scheme 2.13**).



Scheme 2.13: The cleavage of library **L 2.1** from Rink-PEGA resin;
Reagents and conditions: (i) TFA/TIS/DCM (90:5:5), 1h.

In conclusion, a PNA-encoded tetrapeptide library (**L 2.1)** was synthesised using the split-and-mix approach, with a general structure of this library shown in **Figure 2.10**.

The removal of the Alloc group on the N^{ε} of D-Lys of **2.27** (**Scheme 2.12**) was often unsuccessful and problematic (perhaps best described as fickle!). Sometimes, after cleavage of the library from the resin, the library was found not to contain FAM (according to UV-Visible spectroscopy). This lack of labelling was clearly a major concern as it came right at the end of the synthesis (rhodamine was always successfully coupled). Therefore removal of Alloc group was investigated in detail.

Tetrapeptide **2.29** (RHOD-Ahx-L-Ala-D-Ala-D-Ala-D-Ala-D-Lys(Alloc)) was synthesised on PEGA resin and its Alloc deprotection was investigated using combinations of palladium sources, scavengers and reaction times (**Table 2.3**)[28]. Despite the number of conditions tested, no Alloc deprotection of peptide **2.29** was detected.

Palladium source	Scravenger	Reaction time
$Pd(PPh_3)_4$ (0.1 equiv.)	$PhSiH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(PPh_3)_4$ (0.1 equiv.)	$Me_2NH{\cdot}BH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(OAc)_2$ (0.1 equiv.)	$PhSiH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(OAc)_2$ (0.1 equiv.)	$Me_2NH{\cdot}BH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(PPh_3)_4$ (1.0 equiv.)	$PhSiH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(PPh_3)_4$ (1.0 equiv.)	$Me_2NH{\cdot}BH_3$ (5.0 equiv.)	15 min. (x2)
$Pd(PPh_3)_4$ (0.1 equiv.)	$PhSiH_3$ (5.0 equiv.)	30 min. (x2)

Table 2.3: Reaction conditions used in attempts of Alloc deprotection[28].

In conclusion, since the removal of the Alloc group was so problematic this protecting group could not be used reliably in the library synthesis. Another library (**L 2.2**) was designed and synthesised using a different strategy.

2.4.4 Synthesis of a 1296 member D-amino acid-containing peptide library.

In this tetrapeptide library, each amino acid was encoded with a PNA quartet containing three PNA monomers (A, C and T). Each position in the tetrapeptide was varied with 6 different amino acids thus the library consisted of up to 1296 members (see **Table 2.4**).

D-amino acids used in the library synthesis contained a variety of side-chain properties [basic (D-Lys), acid (D-Glu), aromatic (D-Tyr), cyclic (D-Pro), small (D-Ala) and achiral (Gly)].

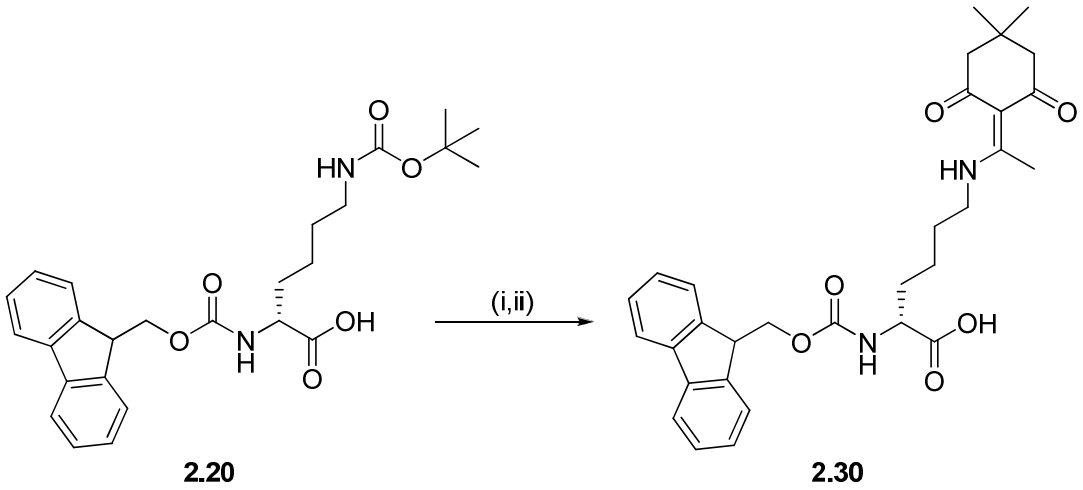
Amino acids	PNA code (*N* to *C*)
D-Lys	TCTA
D-Glu	ATCA
D-Tyr	ACAT
D-Pro	AATC
D-Ala	TACT
Gly	TTAC

Table 2.4: Amino acids and PNA codes used in the split-and-mix library synthesis of libraries **L 2.2** and **L 2.3**.

The library **L 2.2** was designed such that it had FAM at the N-terminal end of the peptide and RHOD at the N-terminal end of the PNA tag. Although FRET might not be obtained from library **L 2.2**, cleaved sequences could be identified from the ratio and relative decreases of FAM intensity. Because of the large number of library members, the variation of hybridisation efficiency or differences in concentration

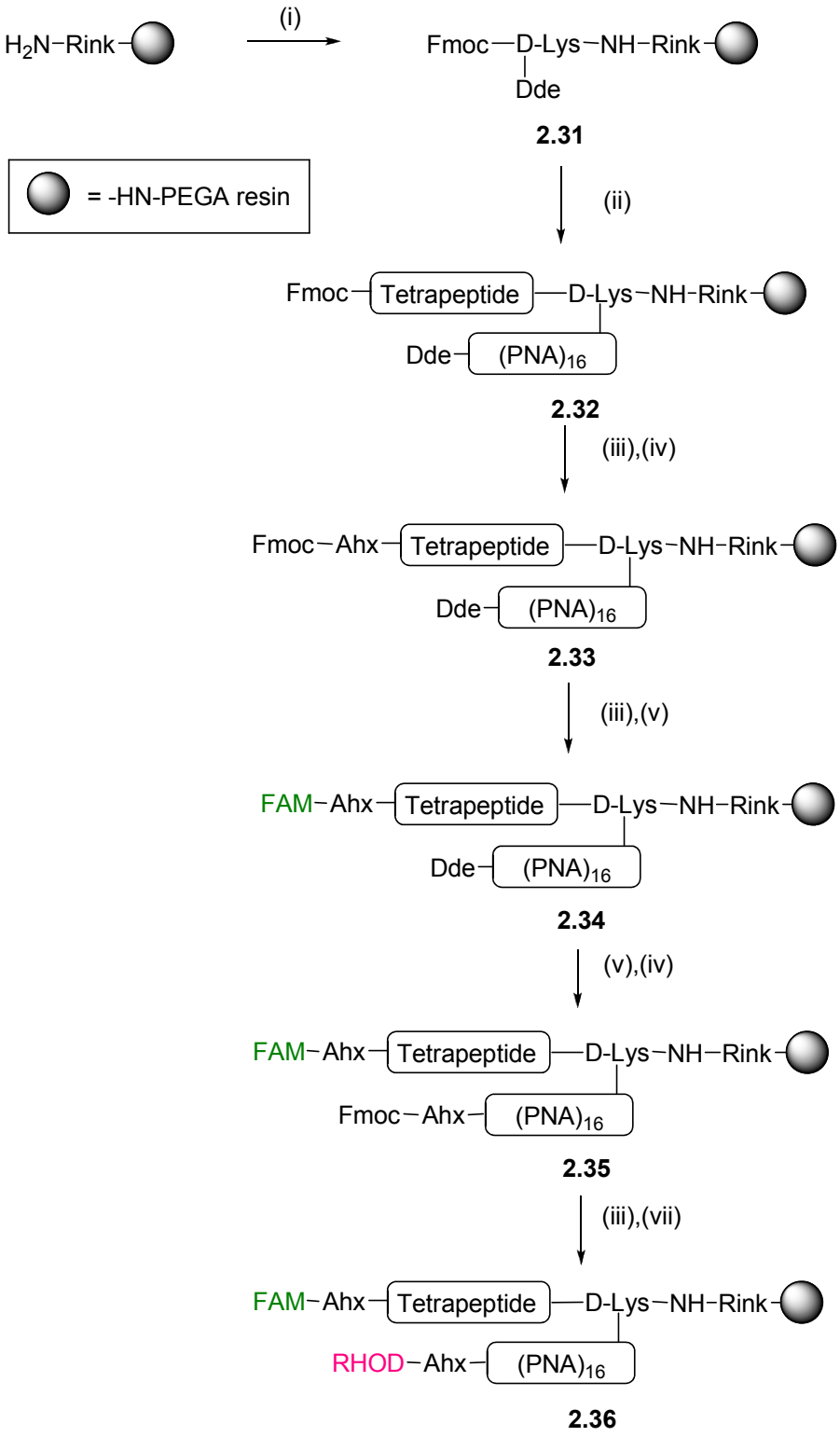
across the microarray might exist; therefore, decreases in FAM intensities might be the result of these causes besides the cleavage of peptides. RHOD at the PNA oligo would thus be an internal control and the ratios of FAM/RHOD would be used instead of absolute FAM intensities.

In library **L 2.2**, PNA tags were conjugated to a tetrapeptide *via* a D-Lys core. Therefore, orthogonally protected, Fmoc-D-Lys(Dde)-OH (**2.30**), was firstly synthesised as shown in **Scheme 2.14**[27].



Scheme 2.14: Synthesis of Fmoc-D-Lys(Dde)-OH (**2.30**); *Reagents and conditions*: (i) 1:1 TFA:DCM, 2 h; (ii) 1.0 equiv. **2.3**, 0.1 equiv. TFA, EtOH, reflux[27].

The synthesis of library **L 2.2** is summarised in **Scheme 2.15**. Synthesis started with the coupling of Fmoc-D-Lys(Dde)-OH (**2.30**) onto H-Rink-PEGA resin. Resin **2.31** was then used for library synthesis (**Table 2.4**). The library (**2.32**) was modified by coupling with the Fmoc-Ahx-OH spacer onto the peptide chain (**2.33**), followed by 5(6)-carboxyfluorescein, giving the FAM labeled PNA-encoded tetrapeptide library (**2.34**). The same spacer, Fmoc-Ahx-OH, was introduced onto the PNA tag after removal of the Dde group, and 5(6)-carboxyrhodamine B was coupled to the Ahx spacer to give **2.36**.



Scheme 2.15: Synthesis of a PNA-encoded tetrapeptide library containing D-amino acids as potential substrates for chymopapain (**L 2.2**); *Reagents and conditions:* (i) **2.30**; (ii)Split-and-mix library synthesis; (iii) 20% piperidine / DMF, 2 x 10 min; (iv) Fmoc-Ahx-OH; (v) 5(6)-carboxyfluorescein; (vi) $NH_2OH{\cdot}HCl$/imidazole in NMP/DMF (5:1), 1 h.; (vii) 5(6)-carboxyrhodamine B; (5.0 equiv. amino acids, PNA monomers or dye, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, microwave irradiation).

The dual colour-labelled library (**2.36**) was cleaved from the solid support and deprotected as in **Scheme 2.13** to give library **L 2.2**, whose structure is shown in **Figure 2.11**, and which was obtained upon precipitation. UV/Vis analysis confirmed the presence of both dyes and their relative equimolar concentrations.

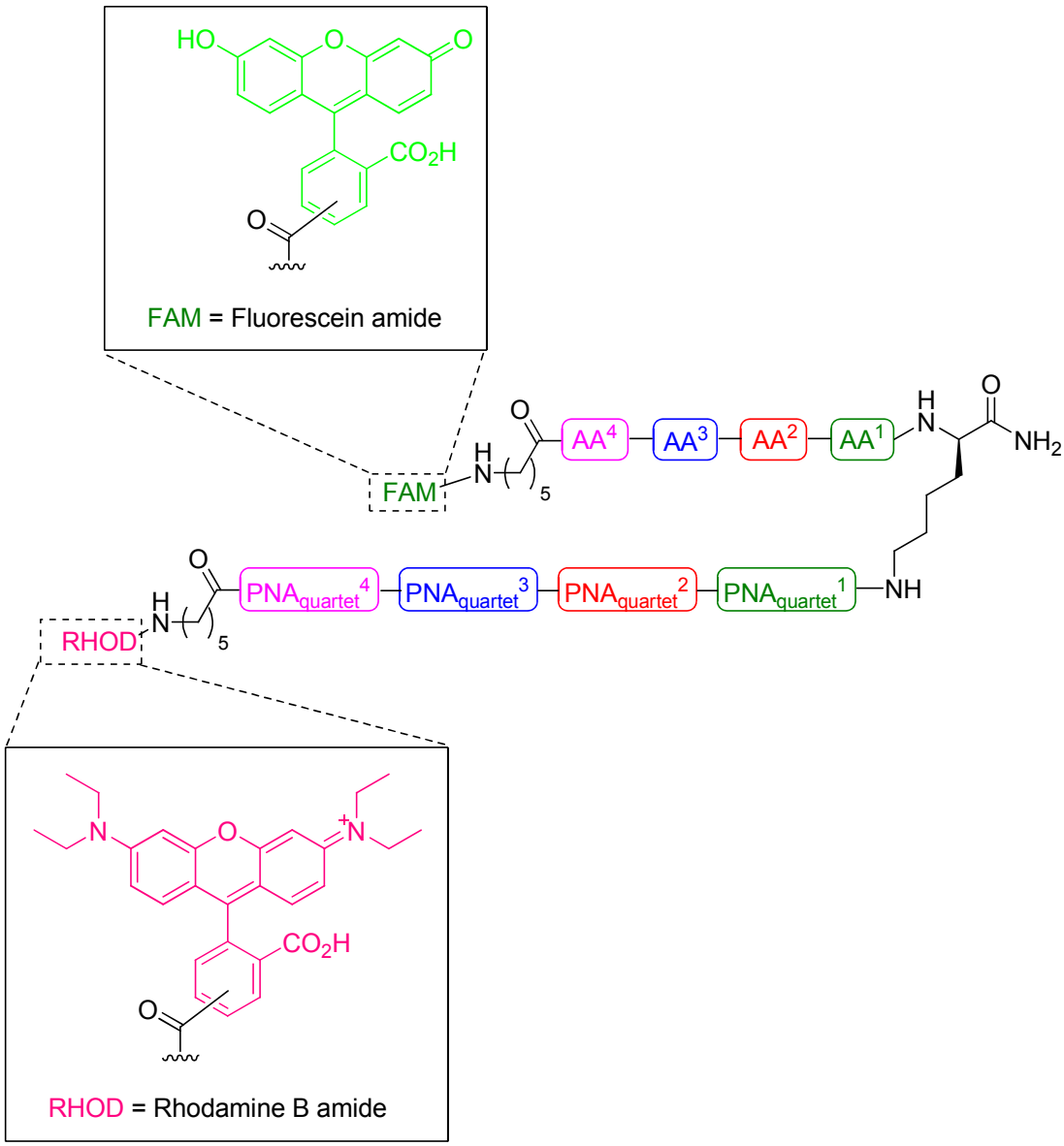
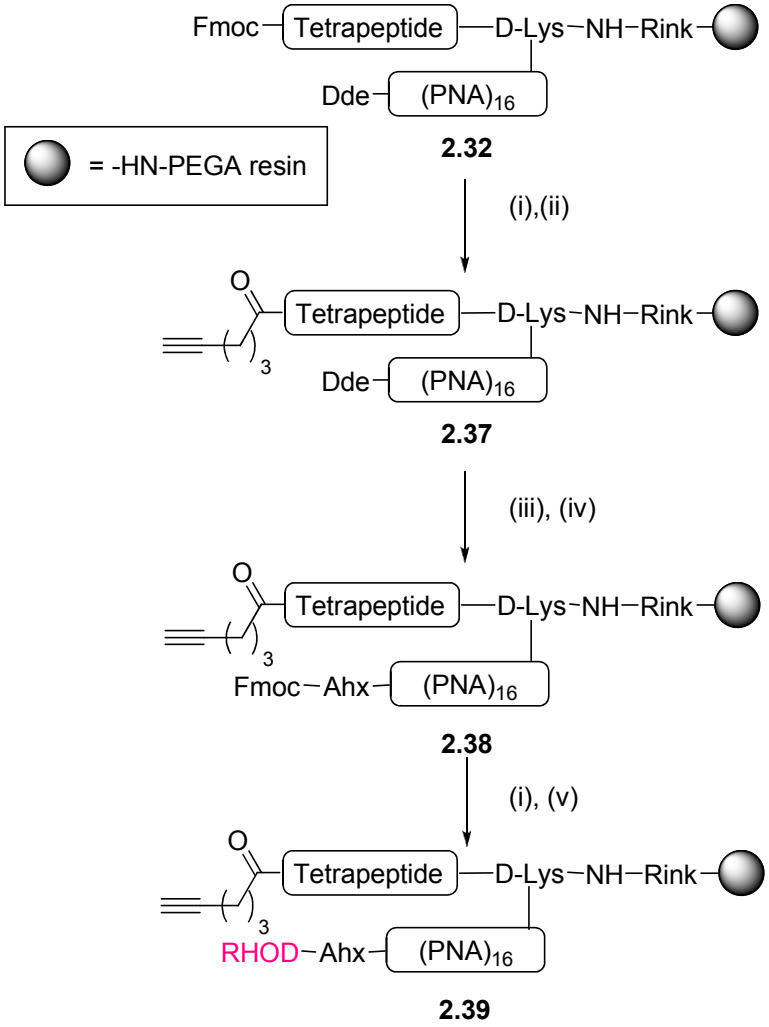


Figure 2.11: Library **L 2.2**.

2.4.5 Synthesis of a 1296 member alkyne-modified peptide library.

In the screening of the peptide library as potential substrates for metal-free Hüisgen cycloaddition (**L 2.3**), an alkyne functional group needed to be available on the peptide. The synthesis of library **L 2.3** is shown in **Scheme 2.16**, starting from **2.32** (**Scheme 2.15**).



Scheme 2.16: Synthesis of PNA-encoded tetrapeptide library consisting of D-amino acid as potential substrates for metal-free Huisgen cycloaddition (**L 2.3**); *Reagents and conditions:* (i) 20% piperidine / DMF, 2 x 10 min; (ii) 5-hexynoic acid; (iii) $NH_2OH{\cdot}HCl$/imidazole in NMP/DMF (5:1), 1 h.; (iv) Fmoc-Ahx-OH; (v) 5(6)-carboxyrhodamine B; (5.0 equiv. amino acids, PNA monomers or dye, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, microwave irradiation).

Resin **2.32** was deprotected, followed by coupling of 5-hexynoic acid (**2.37**). The next step was modification of the PNA chain. Fmoc-Ahx-OH was coupled as a

spacer after removal of the Dde group (**2.38**). After deprotection of the Fmoc group, 5(6)-carboxyrhodamine B was coupled, giving rhodamine B amide-labelled library (**2.39**).

After cleavage of **2.39** with TFA and precipitation, library **L 2.3** (**Figure 2.12**) was obtained as a red solid. The presence of the acetylene group was confirmed by both the qualitative nynhidrin test (before cleavage), which gave the expected negative result following coupling of 5-hexynoic acid (**2.37**) and infrared spectroscopy analysis which showed characterestic peaks for an alkyne at 3273 and 2159 cm^{-1}.

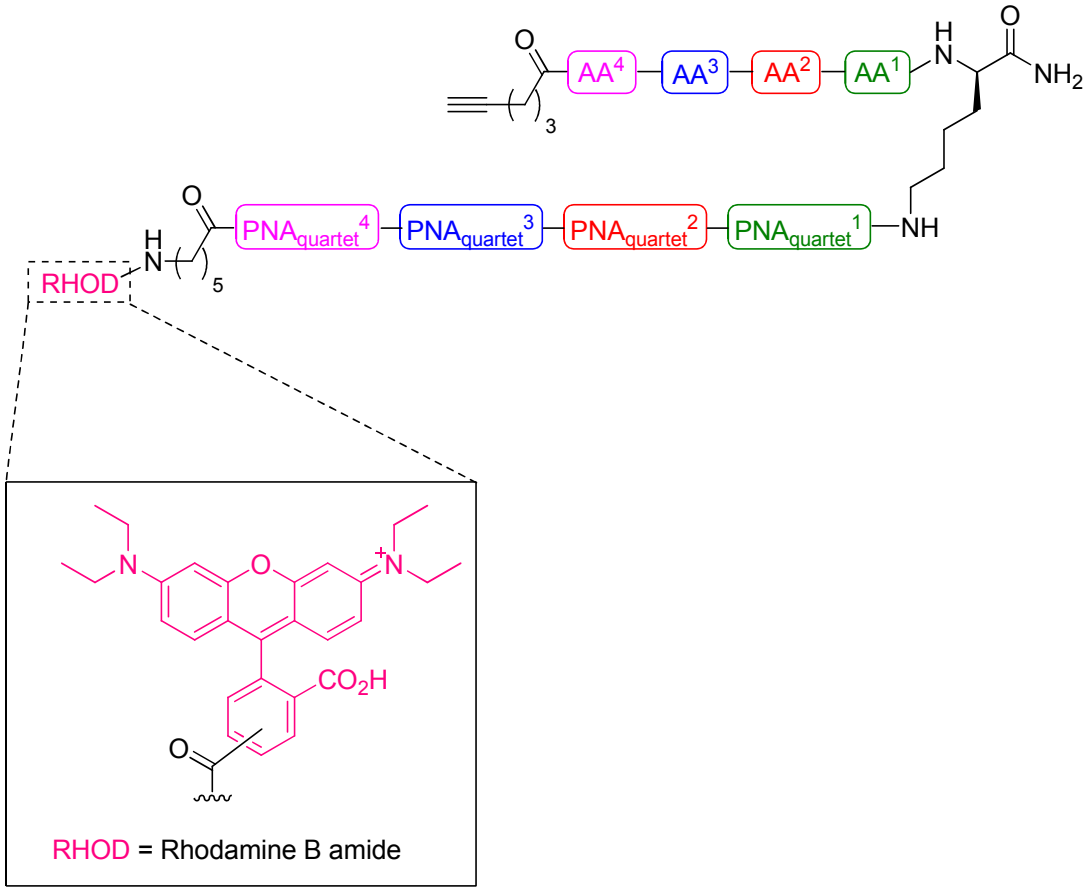


Figure 2.12: Library **L 2.3**.

2.4.6 Synthesis of a 625 member PNA-encoded cell-penetrating peptide library.

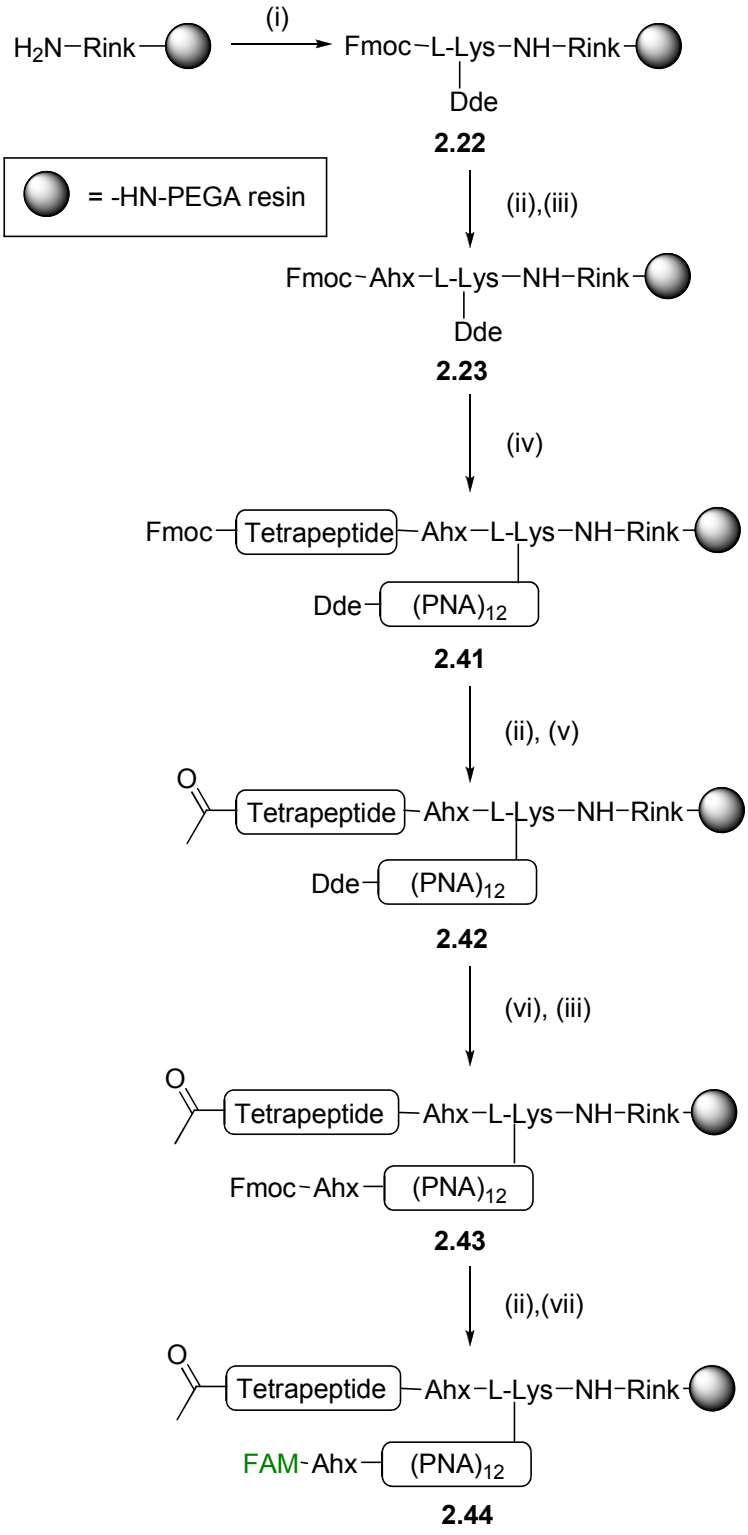
A 625-membered library of PNA encoded cell-penetrating peptide library (**L 2.4**) was synthesised. Each position of the tetrapeptide library was varied with 5 building blocks, each of which was encoded by a unique PNA triplet.

Besides amino acids, *N*-6-aminohexyl-glycine (**2.40**, also known as lysine-like peptoid monomer, Llp) was used as one of the building blocks. The building blocks and PNA triplets used in the synthesis of library **L 2.4** are shown in **Table 2.5**.

Amino acid (AA^1-AA^2)	Amino acid (AA^3-AA^4)	PNA code (*N* to *C*)
2.40	**2.40**	TAA
Glu	Glu	ATA
Pro	Pro	TCC
Gly	Gly	CTC
Cys	Tyr	AAC

Table 2.5: Building blocks and their PNA triplet codes used in split-and-mix library synthesis of cell-penetrating peptide library (**L 2.4**).

The synthesis of **L 2.4** is summarised in **Scheme 2.17**. A PNA-encoded split-and-mix library synthesis was performed starting from **2.22**, on which L-Lys was used as a core of the conjugate, resulting in the library on the solid support **2.41**. After removal of the Fmoc group, the peptide chain was capped by reacting with acetic anhydride (**2.42**). The PNA tag was coupled to the Ahx spacer after Dde deprotection (**2.43**), followed by labelling with 5(6)-carboxyfluorescein to give the FAM-labelled PNA-encoded tetrapeptide library **2.44**. After TFA cleavage of **2.44**, the library **L 2.4** (**Figure 2.14**) was obtained.



Scheme 2.17: Synthesis of a PNA-encoded tetrapeptide library of potential cell-penetrating peptides; *Reagents and conditions:* (i) **2.19**; (ii) 20% piperidine/DMF, 2 x 10 min; (iii) Fmoc-Ahx-OH; (iv)Split-and-mix library synthesis; (v) 1:1 (v/v) Ac_2O:Py, r.t., 1.5 h.; (vi) $NH_2OH{\cdot}HCl$/imidazole in NMP/DMF (5:1), 1 h.; (vii) 5(6)-carboxyfluorescein; (5.0 equiv. amino acids, PNA monomers or dye, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, microwave irradiation).

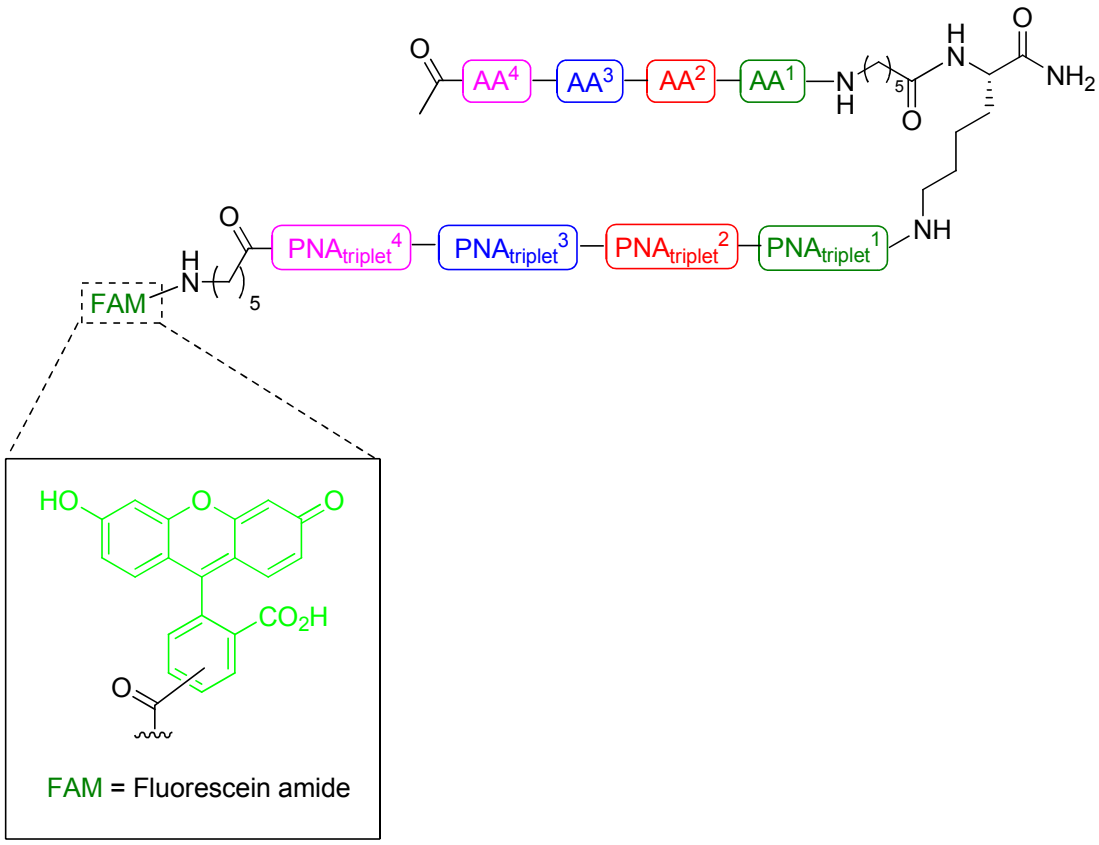


Figure 2.13: Library **L 2.4**.

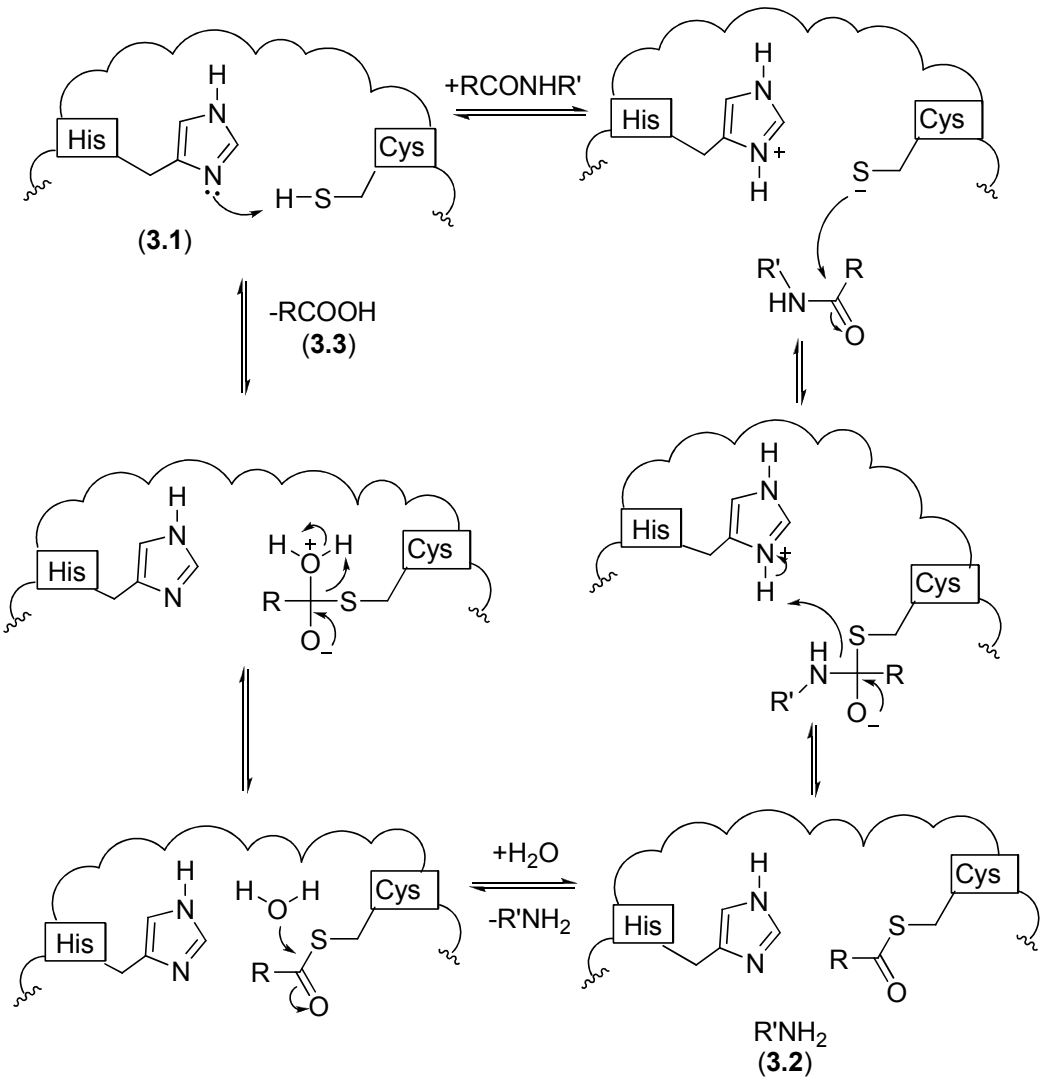
Chapter 3: Screening of an encoded D-amino acid containing peptide library – looking for new and unusual potential protease substrates.

3.1 Screening of a PNA-encoded peptide library with chymopapain.

3.1.1 Chymopapain.

Chymopapain is a cysteine protease, obtained from the plant *Carica papaya*, that is used for a number of biomedical applications. It is used for pain relief in the treatment of herniated invertebral disks (chymodiactin)[29,30], thereby reducing pressure on the nerve root, whereas orally administered chymopapain is used to treat intestine cuticles produced by neumatoids[31]. It is also one of the components of contact lens cleaning liquids[32].

The catalytic mechanism of cysteine protease is summarised in **Scheme 3.1**. The cycle starts with deprotonation of a thiol in the enzyme's active site by a basic side chain, usually a histidine residue (**3.1**). The nucleophilic attack on the peptide's carbonyl carbon by the deprotonated cysteine side chain leads eventually to the release of the amine (**3.2**). The next step is the release of carboxylic acid (**3.3**) (*via* cleavage of the thioester) as well as the recovery of the free enzyme by hydrolysis.



Scheme 3.1: Catalytic cycle of cysteine protease.

3.1.2 Fluorescence resonance energy transfer.

Fluorescence resonance energy transfer is a phenomenon of energy transfer between two chromophores. The donor fluorophore, when excited, transfers energy to the acceptor chromophore, avoiding the release of this energy at its emission wavelength. FRET takes place if these two conditions occur:

1. Both chromophores must be close, typically less than 10 nm.
2. The emission spectrum of the donor chromophore must significantly overlap with the absorption spectrum of the acceptor chromophore.

In an ideal situation, the donor chromophore is excited at λ_1, emits at λ_2 and acceptor chromophore is excited at λ_2, emits at λ_3, (the overall process can be summarized as shown in **Figure 3.1**). In this scenario, when the system is excited at λ_1, emission at λ_3 is observed. In another case, the acceptor chromophore does not emit light and is known as a dark quencher.

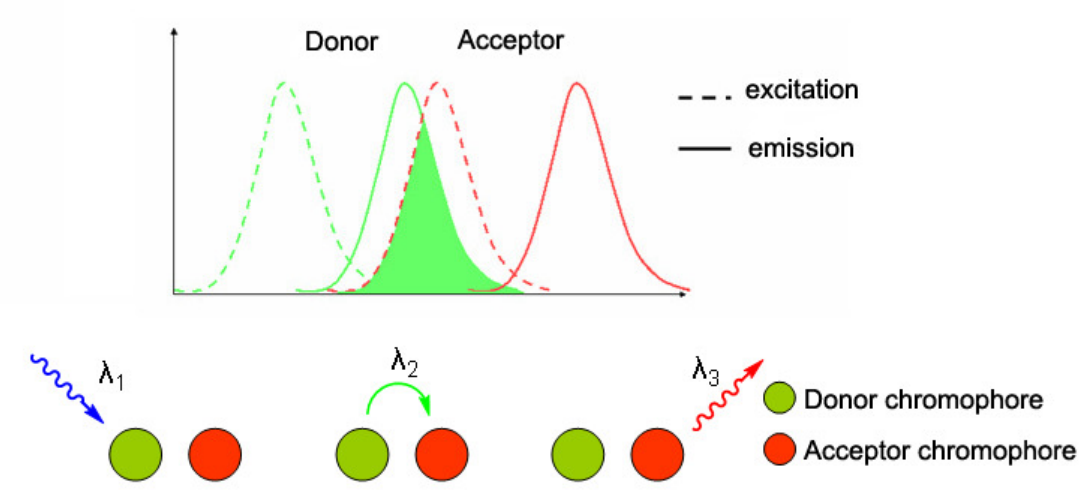


Figure 3.1: Schematic explanation of the FRET principle.

3.1.3 PNA-encoded peptide libraries.

PNA has been used to encode peptide libraries generated using the split-and-mix approach. During the synthesis of the libraries, each building block was encoded (or tagged) with a unique PNA code. The Bradley group has developed a synthetic protocol for the synthesis of PNA encoded peptide libraries, using a set of Dde-protected PNA building blocks[23]. The Dde protecting group is orthogonal to the Fmoc, Alloc and Boc groups, enabling orthogonal library synthesis and tagging. PNA-encoded peptide libraries have been analysed using DNA microarrays with the PNA/DNA hybridisation properties, allowing each and every library member to be analysed by pull-down on a DNA microarray. The libraries consist of a number of variable positions. Each was encoded with a unique PNA triplet. After the incubation

with protease or tyrosine kinase (depending on the purpose of the libraries) each library was hybridised onto a DNA microarray containing DNA sequences complementary to the PNA tags of all library members (with non complementary DNAs as a control).

As an example, a 10,000-member library for profiling tyrosine kinases[16], was developed. Library members were nonapeptides with 4 variable positions, two positions on each side of the Tyr (phosphorylation site) (**Figure 3.2**). Each position was varied with ten amino acids and each amino acid encoded with PNA triplet to give encoding tag composed of 12 PNA monomers. The resulting 10,000-membered peptide library was then treated with Abelson tyrosine kinase, allowing selective members to be phosphorylated.

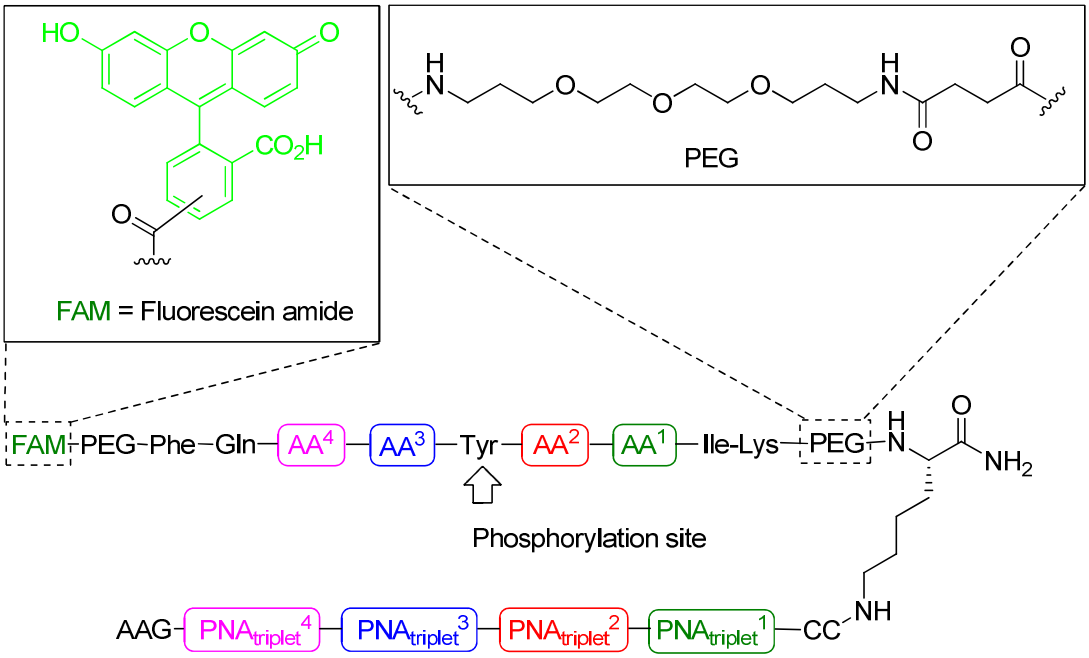
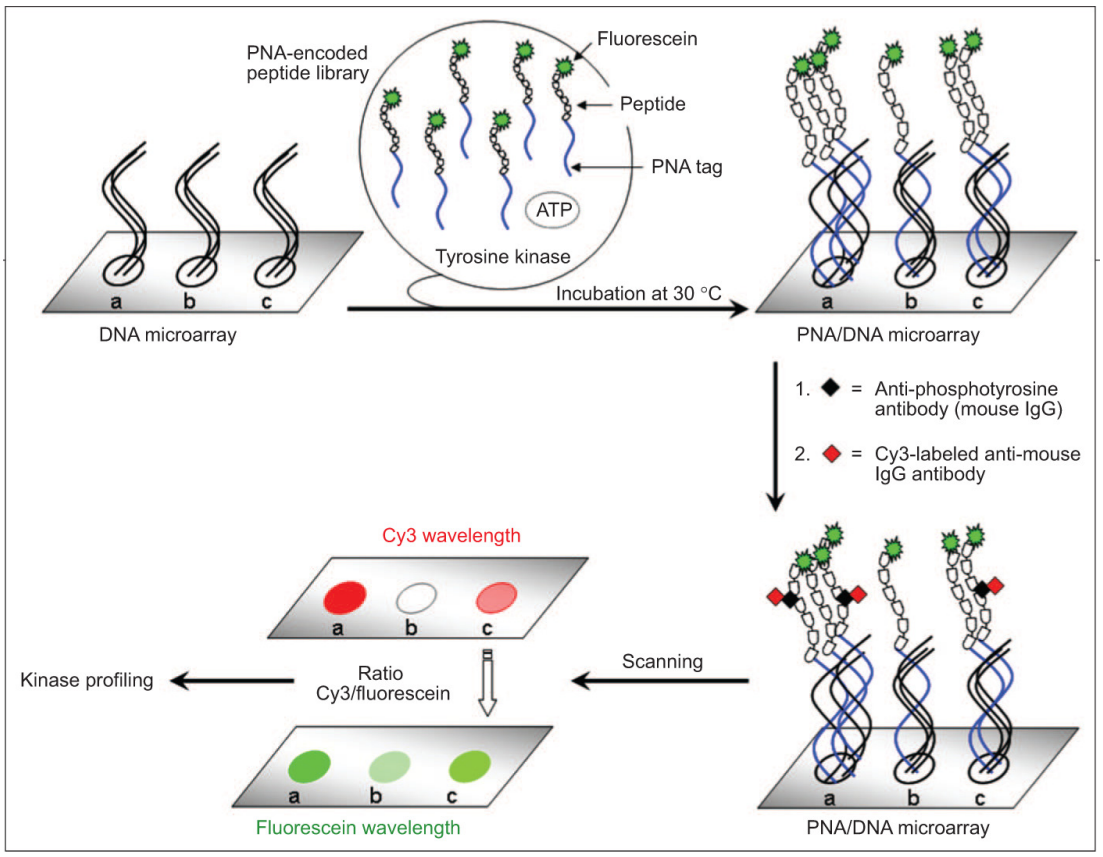


Figure 3.2: Structure of a 10000-membered PNA-encoded library for profiling Tyrosine kinases.

The incubation and hybridisation of the library is shown in **Scheme 3.2**. The library was incubated with Abelson tyrosine kinase (Abl) (the tyrosine kinase that is involved in the regulation of cell proliferation, transcription, and apoptosis) and ATP, resulting in the phosphorylation of any peptide recognized by Abl. The library

was then pulled-down onto a DNA microarray. Each of the 10,000 peptides was delivered to a specific location on the microarray by the formation of antiparallel PNA/DNA duplexes. The arrays used contained 22,575 features with 10,000 oligonucleotide sequences complementary to the PNA sequences (each in duplicate) and 2,575 control DNA sequences that were designed to be non-complementary to any of the PNA tags in the library. The phosphorylated peptide sequences were identified using a primary/secondary antibody approach. An antiphosphotyrosine antibody (mouse IgG) was added onto the array followed by addition of a Cy3-labeled secondary antibody (antimouse IgG), resulting in Cy3 labelling on phosphorylated peptide sequences, which were detected by scanning the microarray using Cy3 filter. Since every library member was labelled with fluorescein, the ratio of Cy3/fluorescein provided an internal control for each and every member of the library, which was crucial due to natural variations in melting temperatures, concentrations, as well as differences in hybridisation efficiency between the 10,000 different library members.



Scheme 3.2: Schematic representation of the hybridisation of a PNA-encoded library incubated with the protein tyrosine kinase Abl and ATP onto a DNA microarray followed by identification of the "hits" using an antiphosphotyrosine antibody and a secondary Cy3-labeled antibody[16].

155 Peptide sequences showing the highest Cy3/fluorescein ratio were identified as "hits". Analysis of these "hit" sequences (**Figure 3.3**) showed that the positions AA^1 and AA^4 have broad amino acid acceptance while AA^2 and AA^3 (the position next to phosphorylation site) have higher specificity of amino acids. Glu was the most frequent amino acid found in position AA^3. In addition, Ser and Val were also observed at this position. For the position AA^2, polar amino acid Ser and acidic amino acid Glu were observed as preferred amino acids.

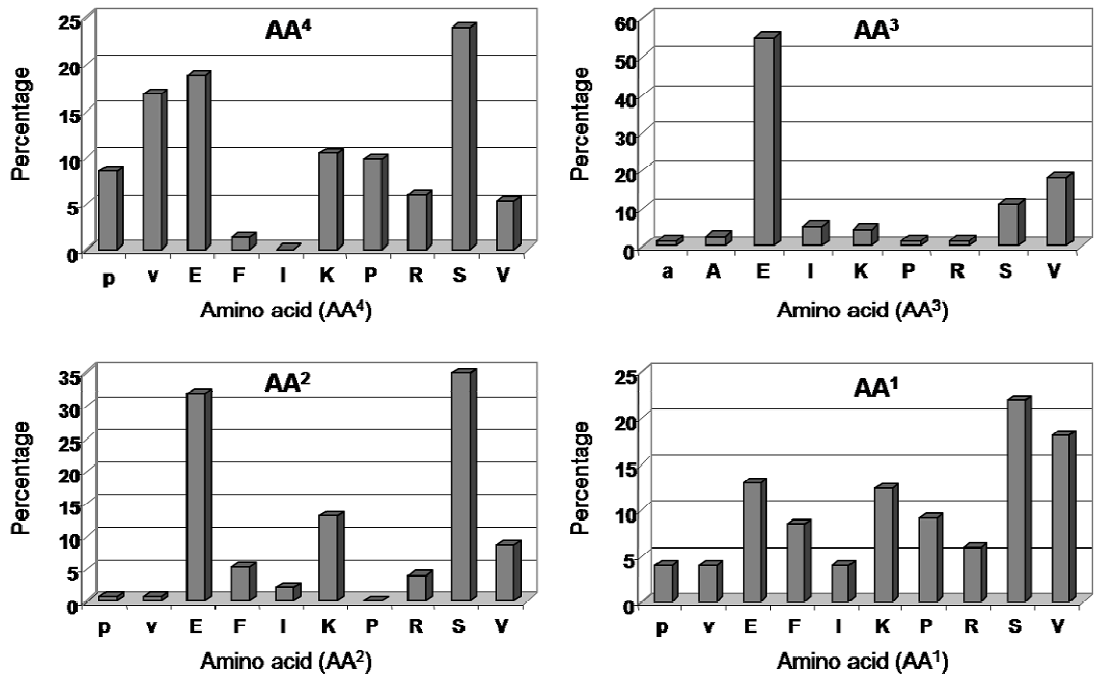


Figure 3.3: Bar graphs for Abl showing the proportion of amino acids at the four randomized positions (AA^4, AA^3,AA^2, and AA^1) in the 155 "hits".

The 10 peptides best recognized by Abl, showing the highest Cy3/fluorescein ratios, are shown in **Table 3.1**. By running a protein basic local alignment search tool (BLAST) search using the Swiss-Prot database using the top 25 sequences from the library, the results showed that the "hit" sequences from the library were in agreement with known substrates for Abl. For example, the second top sequence (SEYES) was the well known Abl substrate c-Jun aminoterminal kinase. These results confirmed the validity of the protocol.

1	SEYEV
2	SEYES
3	SEYSF
4	SEYVF
5	VIYES
6	SIYEP
7	SIYSP
8	PEYSE
9	SEYEE
10	VEYES

Table 3.1: 10 Peptide sequences identified from the 10,000 library screen with Abl kinase.

3.1.4 Screening of a 10,000-membered PNA-encoded peptide library with chymopapain.[15]

A library of FRET-based hexapeptides consisted of four variable positions was synthesised using a PNA-encoded split-and-mix protocol. Each position was varied with 10 amino acids, giving rise to 10,000-membered library (**Figure 3.4**). Each amino acid was encoded with PNA triplet, which had been designed in such a way that the resulting oligomers could be hybridised only in one orientation (PNA *N*-teminus facing the DNA 3'-end) by the introduction of common PNA monomers onto both ends of the oligomers (AAG on the *N*-terminus and CC on the *C*-terminus). These common PNA monomers also prevented self complimentaries and dimer formation.

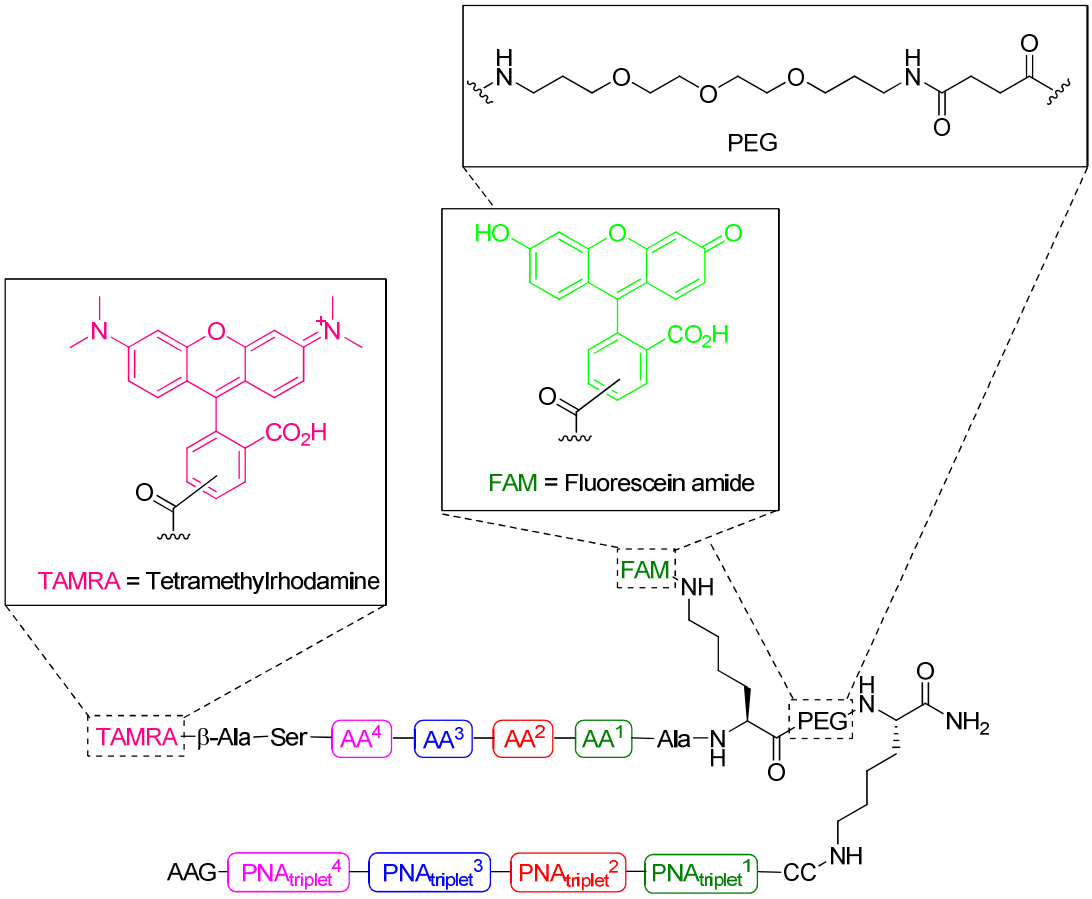


Figure 3.4: General structure of the 10000 member PNA-encoded FRET base peptide library.

The absorption/emission spectra of chromophores used in this library are shown in **Figure 3.5**. The overlap of the fluorescein (FAM) emission wavelength and tetramethylrhodamine (TAMRA) absorption wavelength shows that FRET will take between these chromophores (red area).

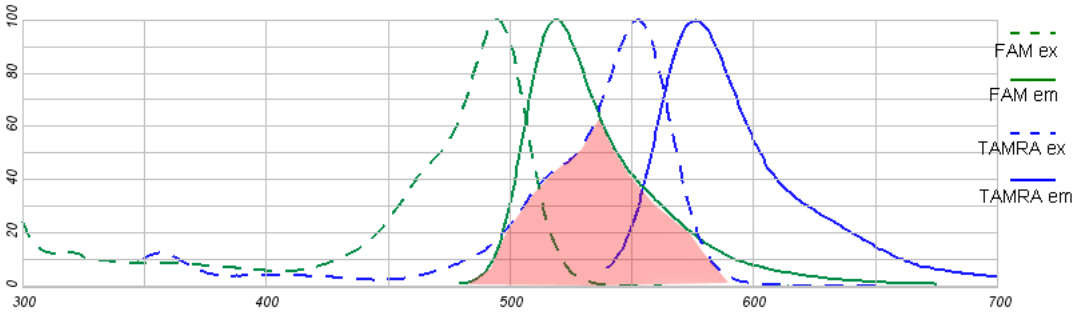


Figure 3.5: absorption/emission spectra of fluorescein and tetramethylrhodamine.

This library was used in the determination of the substrate specificity of chymopapain using a DNA microarray consisting of DNA oligomers complimentary

to all PNA oligos existing in the library (in duplicate). Upon hybridisation, the PNA tags brought each library member to specific position on the DNA microarray having a complementary DNA sequence. Since the DNA sequences of every position on the microarray are known, the peptide sequence of the library member hybridised onto each position could be identified due to the amino acid-PNA tag correspondence.

The library, before treatment with chymopapain, was first hybridised onto a DNA microarray and scanned using TAMRA and FAM filter sets. Analysis of the resulting images showed no signal on the control spots, indicating that no unselective binding occurred in the hybridisation process while the presence of fluorescent intensities confirmed the success of library synthesis.

After treatment with chymopapain, the library was hybridised onto a DNA microarray and the array was scanned similarly to that of untreated library. The ratios of the FAM/TAMRA intensities were analysed and for the "hit" sequences, the FAM/TAMRA ratios would be expected to be significantly increased due to the cleavage of TAMRA from the peptide (**Figure 3.6**).

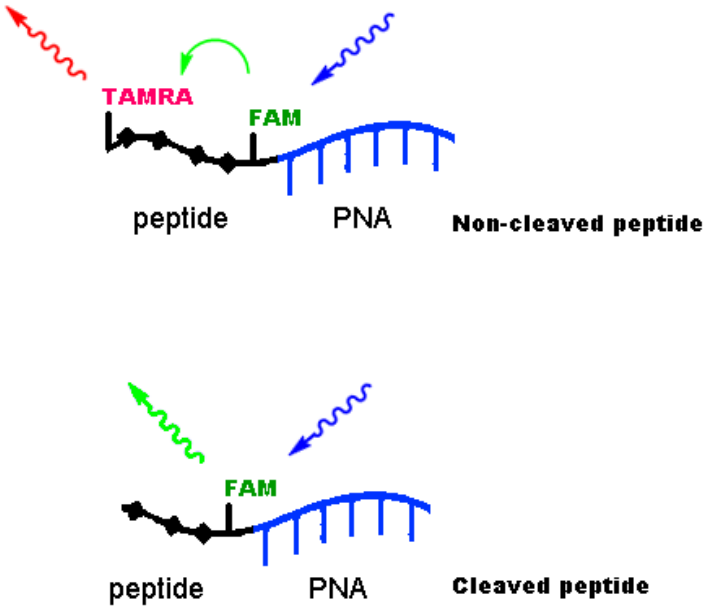


Figure 3.6: Comparison of cleaved and non-cleaved peptide sequences.

3.1.5 Analysis and identification of possible "hit" peptide sequences derived from the screening of the 10,000-member PNA-encoded peptide library.

The 10,000-membered PNA encoded peptide library contains FRET pair FAM and TAMRA. Besides being a FRET acceptor, TAMRA also acted as an internal control. The internal control was necessary in the library screen since the factors that affect fluorescent intensities exist, including a broad range of melting temperature, variation in hybridisation efficiency and difference in concentrations of library members across the array. The fluorescent ratios were used instead of absolute fluorescent intensities in order to minimize these effects. Therefore the cleaved peptide sequences were identified by the high ratios of FAM (of treated library) to TAMRA (of control library).

Unmodified library was first hybridised on the DNA microarray as a control. The customized microarray contained 22,575 spots (10000 DNA sequences complementary to PNA tags in duplicate plus 2,575 non-complementary sequences). Microarray scanning (using both TAMRA and FAM filter sets) revealed that the fluorescent signals could be detected only on spots complementary to PNA tags and were not homogenous over the array (**Figure 3.7**).

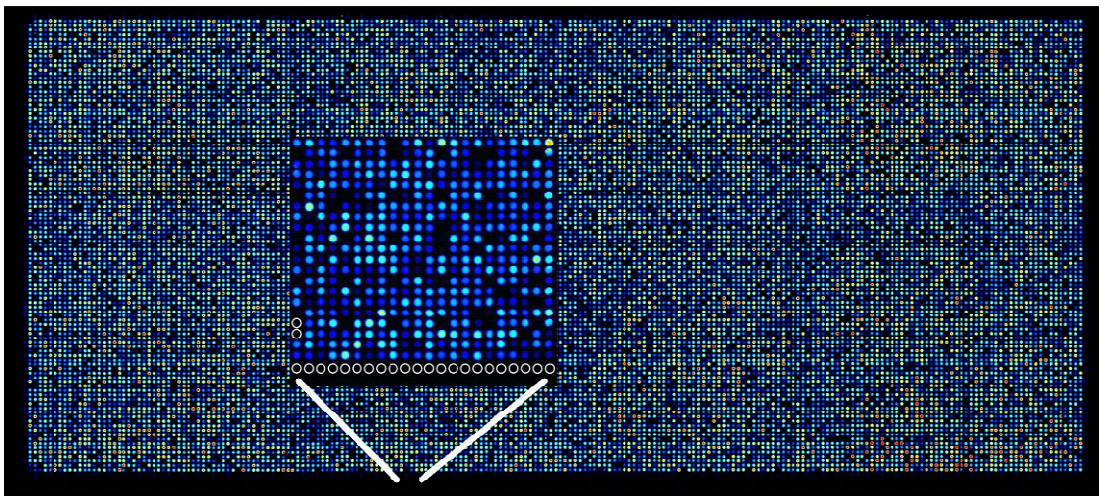


Figure 3.7: A 10 000 member uncleaved PNA-encoded library hybridised onto a 22575 custom DNA microarray. Empty circles represent spots containing DNA oligomers that were not complementary to any PNA oligomer in the library.

Following enzymatic treatment, the library was hybridised onto the DNA microarray and scanned. The FAM/TAMRA ratios were obtained and accepted if the standard deviation of the duplicates were less than 0.25, resulting in 8,685 acceptable data points (**Figure 3.8**). From the 8,685 peptide sequences the 100 sequences showing the highest FAM/TAMRA ratios (**Table 3.2**) were picked for further analysis.

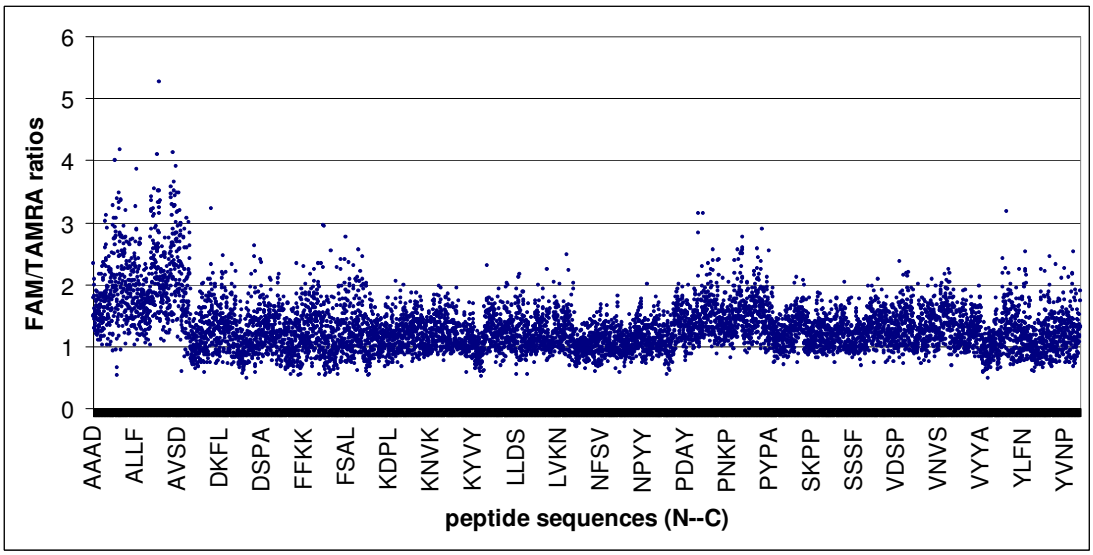


Figure 3.8: Peptide sequences whose FAM/TAMRA ratios have standard deviation of the duplicate less than 0.25.

APVV	AFNK	AKKY	FLLL	AVPA
AFSV	AFVV	AVSL	AFDF	AVNL
AVFK	AFVY	YFLS	AVYK	PFKV
APPV	APYV	APAF	AKFP	APVP
AFKF	APAV	AVNS	APNV	AFVS
AVLY	APVK	PFLK	AVSF	AVFL
ALLY	APVL	APVN	FLLV	AFVL
AVFS	AFSL	PFVS	AFKD	AVLF
AVAF	AVAV	APKF	APLF	AVYF
APKV	AVKS	ADKV	ADLY	AFVK
APVS	AVFY	AVDK	APAS	AVVK
AVKF	AFDK	APFV	PVVL	AFKS
APVA	ALLL	AFKY	ALLS	AFYS
AVVF	AVFV	AYPV	AVFN	AYKF
AVPK	APVY	AFFK	AYKV	PPVK
AFPV	AFSF	ADKK	ALLV	FPVV
AVAL	APFK	AVNF	AVYS	APAA
AVKY	DFKS	AYVK	APVD	AFYK
APAL	APAK	AVYV	AFLK	AKLL
AVAK	AVVY	AVYL	AYYK	AYKD

Table 3.2: The 100 peptide sequences (N - C) showing highest FAM/TAMRA ratios.

The preference amino acids in each position were identified using the data in **Table 3.2** (**Figure 3.9**). Chymopapain showed the highest specificity for Ala at position AA^4, Phe/Pro/Val at position AA^3 and Phe/Lys/Ser/Val at position AA^1 while there was little preference for any amino acid at position AA^2. In conclusion, the "hit" sequences from the screening of PNA-encoded library as substrates for chymopapain were Ala-(Phe/Pro/Val)-Xxx-(Phe/Lys/Ser/Val), where Xxx represents any amino acid.

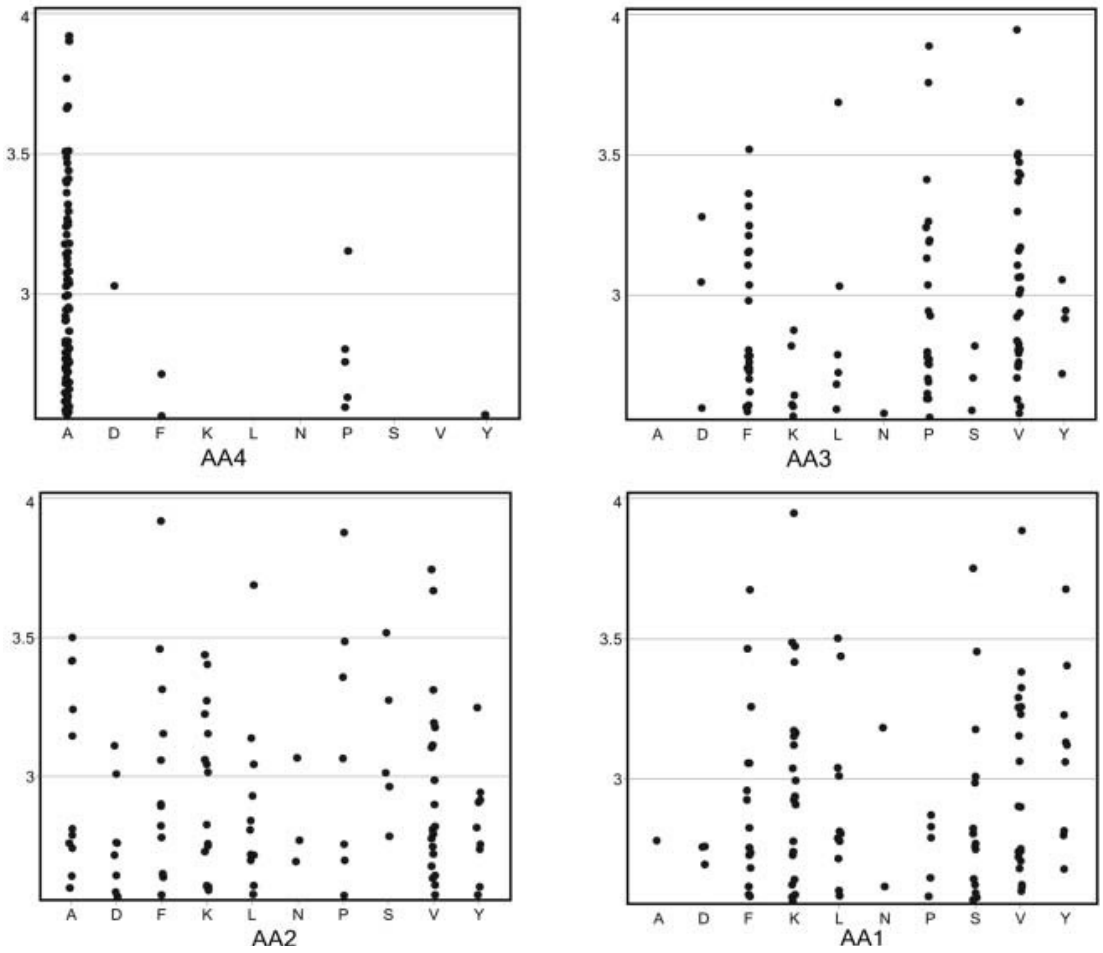
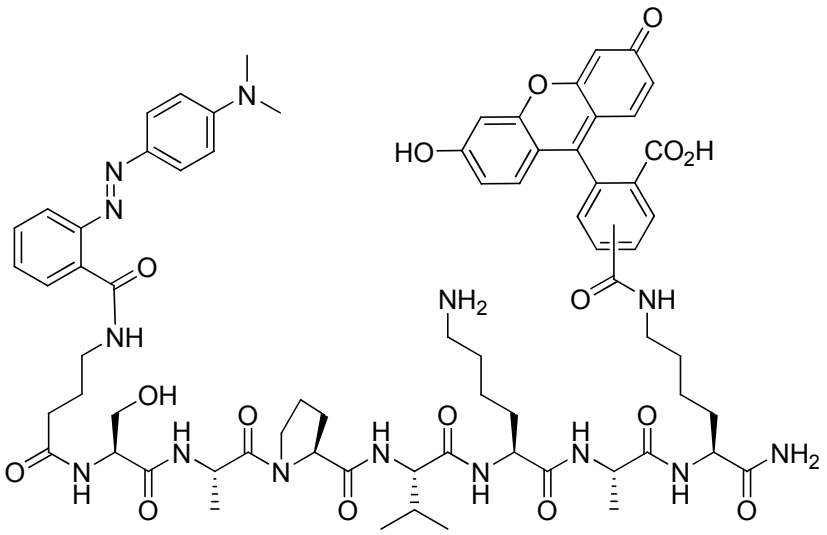


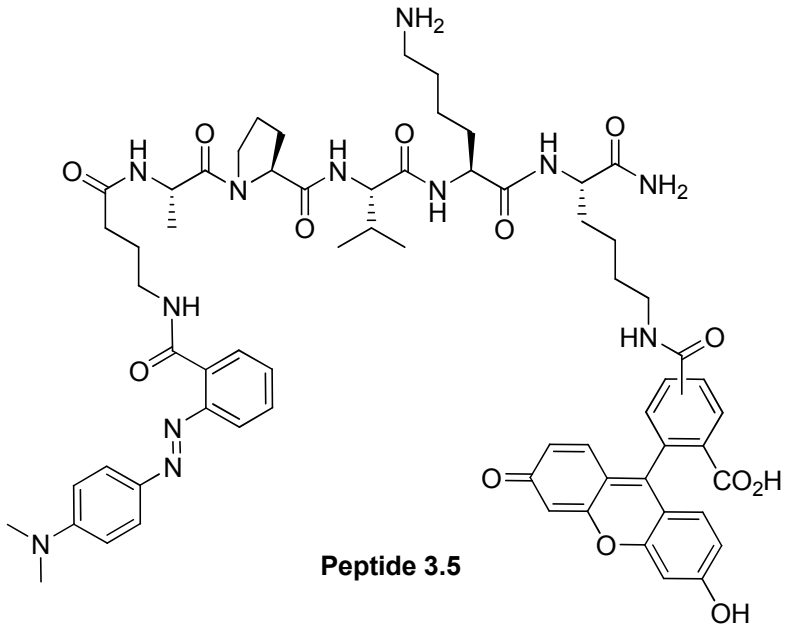
Figure 3.9: Representation of the top 100 peptide cleaved by chymopapain from the 10000-membered library.

3.1.6 Enzymatic evaluation of "hit" peptide sequences as substrates for chymopapain and papain.[33]

Two peptides sequences were designed and synthesised based on the generated "hit" sequences and evaluated as potential chymopapain substrates. These sequences were the hexapeptide Ser-Ala-Pro-Val-Lys-Ala and the tetrapeptide Ala-Pro-Val-Lys. A FRET-pair was used in order to follow the cleavage of substrates (fluorescein (a donor chromophore) and methyl red (a dark quencher)). Fluorescein and methyl red were coupled to the peptides *via* the Lys side chain and a γ-aminobutyric acid spacer respectively (**Figure 3.10**).



Peptide 3.4



Peptide 3.5

Figure 3.10: The FRET peptides used in the enzymatic evaluation of chymopapain; Peptide **3.4**: (Methyl Red-Abu-Ser-Ala-Pro-Val-Lys-Ala-Lys(Fluorescein)-NH_2); Peptide **3.5**: (Methyl Red-Abu-Ala-Pro-Val-Lys-Lys(Fluorescein)-NH_2).

The cleavage of peptides **3.4** and **3.5** was followed using a fluorescence plate reader equipped with 480/20 nm excitation and 520/20 nm emission filters. The experiments were carried out at pH 6.2 and pH 7.4 since chymopapain has a broad working pH range. Enzyme concentrations were determined using λ_{max} = 280 nm and a mass extinction coefficient ($E^{1\%}$) = 18.7[34].

Stock solutions (10 mM) of both FRET peptides were prepared in dimethylsulfoxide (DMSO). Before assay, the peptides were diluted into assay buffers (*assay buffer 1* - 0.1 M phosphate-buffered saline [PBS], pH 7.4; *assay buffer 2* - 10 mM sodium acetate, 0.1 mM ethylenediaminetetraacetic acid [EDTA], and 1 mM cysteine, pH 6.2) at a range of concentrations, keeping the DMSO level below 3.5%. Chymopapain (0.3 µM in 10 mM sodium acetate, 0.1 mM EDTA, and 1 mM cysteine, pH 6.2) was added to the reaction mixture to a final concentration of 30 nM. Enzymatic reactions were carried out at 37°C, and the initial rates were calculated using the Biotek Gen5 data analysis software (only the hydrolysis of peptide **3.5** is demonstrated below). The fluorescence profiles following the hydrolysis of peptide **3.5** (at different concentrations of substrate) by chymopapain were obtained (**Figure 3.11)** and the initial rate obtained from the slope of this fluorescence profile.

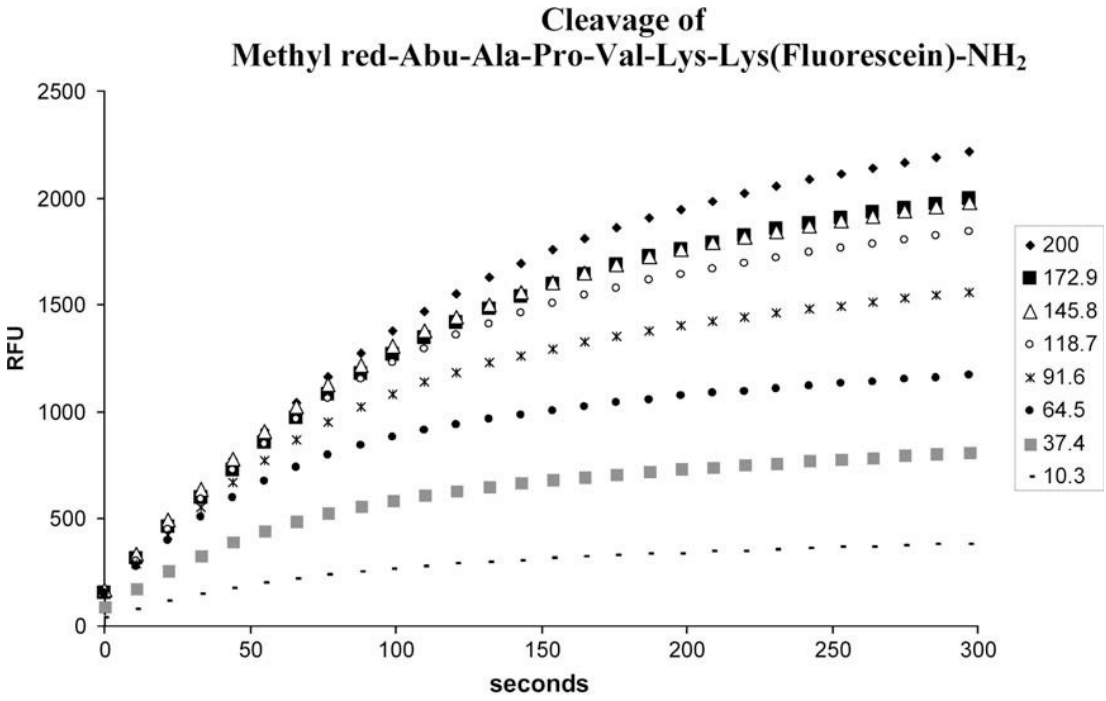


Figure 3.11: Fluorescence profiles (relative fluorescence units [RFU] vs. time [s]) following the hydrolysis of peptide **3.5**, at different concentration of substrate (µM).

From the initial rates obtained from **Figure 3.11**, the initial rates versus substrate concentrations is shown in **Figure 3.12** and this data was used to calculate K_m and V_{max} using the software GraFit.

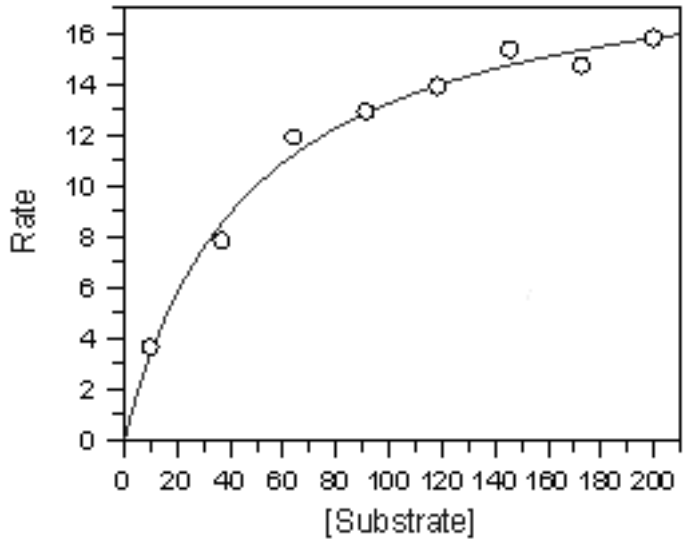


Figure 3.12: Initial rate versus substrate concentration from the cleavage of peptide **3.5**.

The cleavage of peptides **3.4** and **3.5** were analysed by analytical HPLC. **Figure 3.13** shows the two HPLC traces (recorded at 440 nm) of peptide **3.5** before (trace a) and after (trace b) enzymatic incubation, showing complete hydrolysis after 10 min, showing cleavage between Lys-Ala according to MS spectrometry (data not shown).

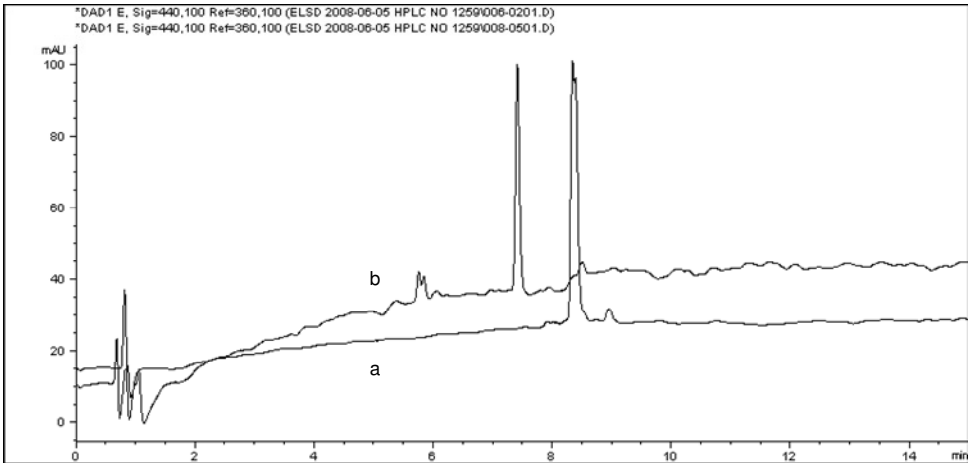


Figure 3.13: HPLC traces recorded at 440 nm of peptide **3.5** before (a) and after (b) treatment with 30 nM chymopapain at pH 7.4 and 37°C.

Kinetic parameters for each peptide are summarized in **Table 3.3**. Both substrates showed K_m values in the low micromolar range. Hydrolysis at different pH showed that pH did not influence K_m. For both substrates, k_{cat} at pH 7.4 was twice as fast as at pH 6.2. The K_m of peptide **3.4** was comparable to that of the commercially available substrates such as benzyloxycarbonyl-lysine phenyl ester (Z-Lys-ONp) (K_m values 7.5-10 μM for chymopapain)[35], but with improved k_{cat}, while specificity would be expected to be superior.

Substrate	pH	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)
Peptide **3.4**	6.2	23 ±7	133 ±1.2	5797
Peptide **3.4**	7.4	26 ±10	360 ±0.3	13,845
Peptide **3.5**	6.2	50.6 ±12	246 ±0.5	4874
Peptide **3.5**	7.4	50.2 ±18	590 ±1.7	11,767

Table 3.3: Kinetic values of the FRET peptides with chymopapain

Papain was also analysed with the same peptides and its results are shown in **Table 3.4**. The results showed that, with increasing pH, papain's affinity for both substrates decreased while the turnover number increased dramatically.

Substrate	pH	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($mM^{-1}s^{-1}$)
Peptide **3.4**	6.2	27 ± 10	243 ± 22	9000
Peptide **3.4**	7.4	101.5 ± 30	1503 ± 5	14,811
Peptide **3.5**	6.2	119 ± 5	475 ± 10	3992
Peptide **3.5**	7.4	200 ± 18	1773 ± 3	6250

Table 3.4: Kinetic values of the FRET peptides with papain.

By comparing the results obtained from chymopapain and papain, it could be summarised as follows:

1. For chymopapain, changing pH between pH 6.2 and pH 7.4 does not affect its affinity against peptide **3.4** and peptide **3.5**. However, papain showed decreased affinity against both peptides with increasing pH.
2. Papain has a higher turnover number than chymopapain for both substrates and both pH's.
3. Although chymopapain and papain have a broad working pH, k_{cat}/K_m value suggested that both enzymes have higher efficiency at pH 7.4.

The scissile bond of substrates was identified by analysing the chymopapain digested samples using liquid chromatography-mass spectrometry. Peptide **3.4** was cleaved at the Lys-Ala bond and peptide **3.5** was cleaved at the Lys-Lys site.

This result showed that lysine was preferable at position P_1 but there was not high specificity for residues in position $P_{1'}$. However, the effect of residues in positions $P_2 – P_4$ could not be concluded since both peptides had identical residues in these positions (**Figure 3.10**). Another peptide having the sequence of Methyl Red-Abu-Ser-Ala-Phe-Lys-Tyr-Ala-Lys(FAM)-NH_2 (peptide **3.6**, **Figure 3.14**) was synthesised and incubated with chymopapain. Peptide **3.6** was found not to be cleaved, suggesting that Val, Pro and Ala at the position $P_2 – P_4$ are important for the target peptides, therefore, the optimal substrate specificity for chymopapain at P_4 to P_1 were found to be Ala, Pro Val and Lys respectively.

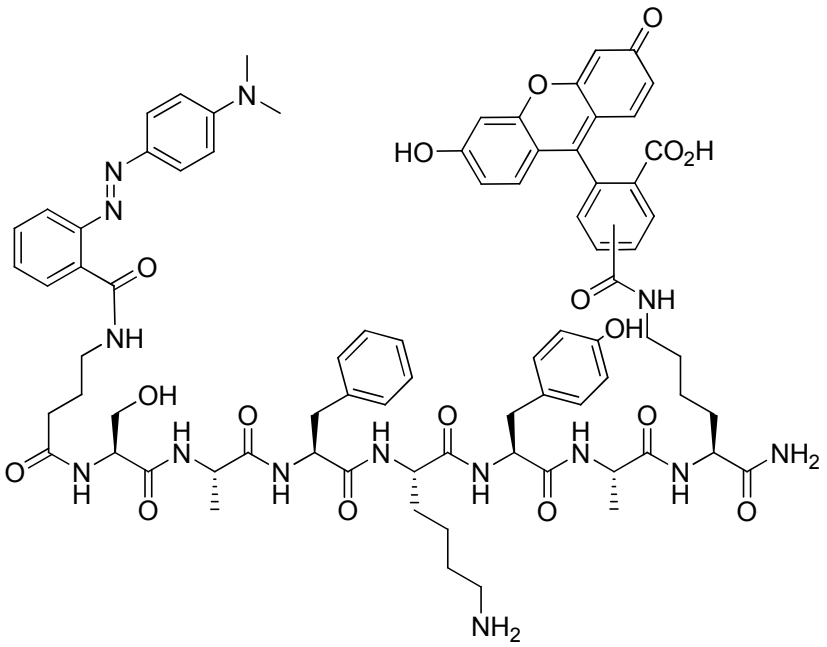


Figure 3.14: Peptide **3.6** (Methyl Red-Abu-Ser-Ala-Phe-Lys-Tyr-Ala-Lys(Fluorescein)-NH_2).

In conclusion, two peptide substrates for chymopapain have been presented for the first time. These peptides showed micromolar affinity toward chymopapain and good turnover numbers, making them perfect candidates for chymopapain studies. In addition, these results also confirmed the success of high-throughput screening of PNA-encoded peptide libraries for the discovery of new substrates for proteases.

3.2 D-amino acid-containing peptides and origins.

In most living organisms, only L-amino acids have been selected as building blocks for peptides and proteins[36], with the D-enantiomer excluded. However, D-amino acids have been found in small amounts in some vertebrates and are widely found in micro-organisms especially in cell wall components.

3.2.1 From bacteria.

The largest source of D-amino acids in nature may be as building blocks of the peptidoglycan bacterial cell wall. Peptidoglycan is a natural polymer presents in

both gram positive and gram negative bacteria. It consists of *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc) sugar residues (**Figure 3.15**)[37], with the MurNAc attached to a pentapeptide which is cross-linked to adjacent peptide strands thereby strengthening the structure. The pentapeptide composition may vary depending on the species of bacteria, but D-glutamic acid and D-alanine are commonly found[38].

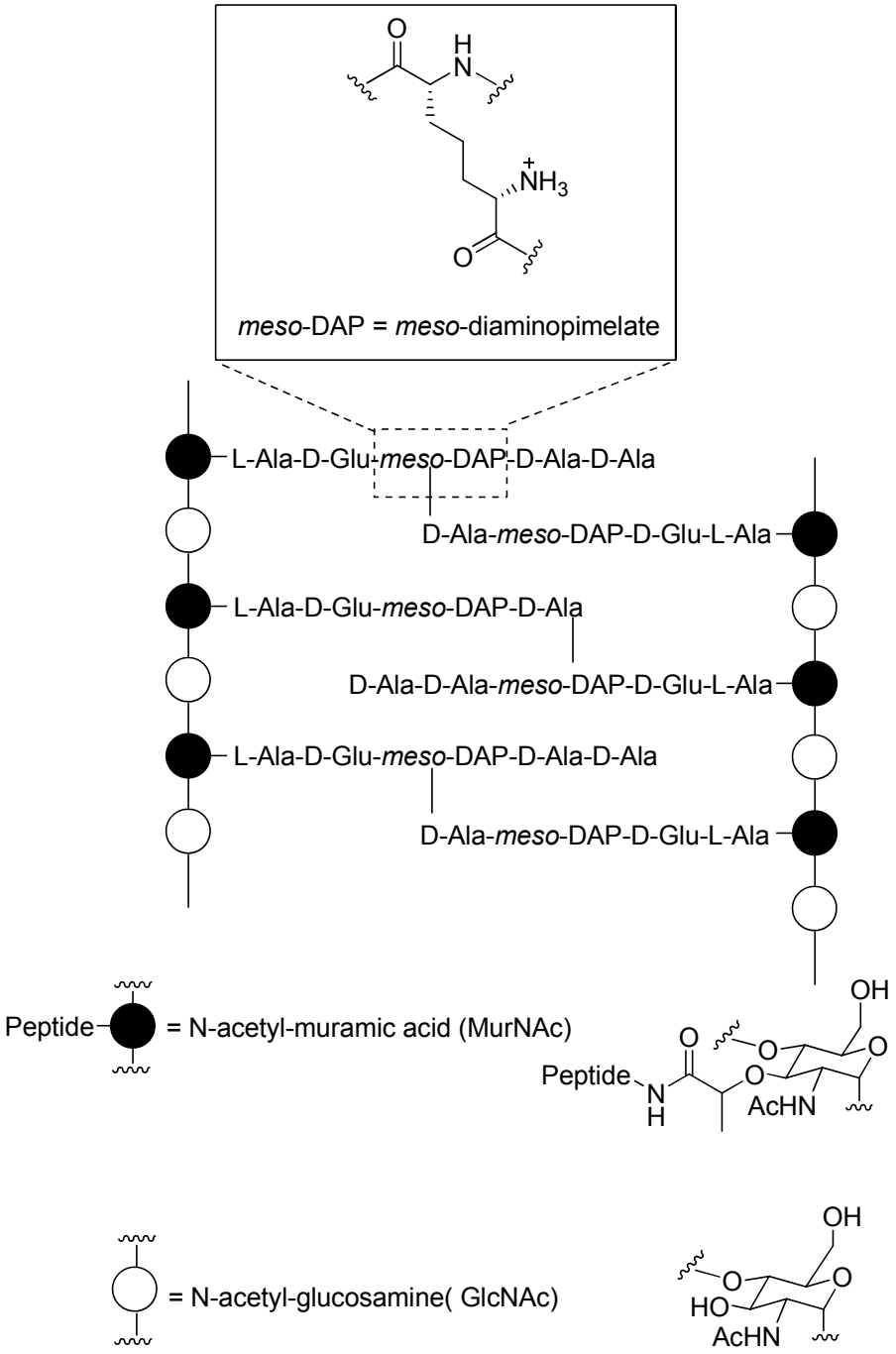


Figure 3.15: Peptidoglycan structure with insert showing structure of *meso*-DAP.

Since peptidoglycan plays a major role in bacteria cell wall integrity, its defect results in cell lysis because of the continuing process of osmosis and the resulting osmotic pressure within the cell causing rupture. As a result, peptidoglycan biosynthesis is a key target for naturally occurring antibiotics. Examples of which are the β-lactam antibiotics, which inhibit cross-linking of the pentapeptides.

3.2.2 Found in humans.

In various human tissues, D-aspartic acid has been detected. In tissues such as teeth[39,40], brain[41-43], and eye lenses[44,45] from the elderly. This presence is believed to be the result of accumulation over time of epimerised L-aspartic acid. There is also a relationship between the presence of this D-amino acid and disease. Thus Roher[43] reported that D-seryl and D-aspartyl residues were found in high levels in brain tissue from people who had died of Alzhiemers disease. Fujii[46] discovered an increase in the D-aspartyl residue at specific positions in a protein which is a component of the eye lens in older humans, with the D-asparatyl level related to levels of cataract disease.

In addition to being part of some peptides, D-amino acids are also found in free-acid form in mammalian tissues. D-aspartic acid has been found in rat brain, the retina's of chicken's, rat's and human's in the early development period although they quickly decrease to the same amount as in adult tissue by the time of birth[36]. Free D-serine is also detected in many areas of the human brain, with the D-serine level very high during the embryonic period. In the frontal cortex, this level stays throughout adulthood[47].

3.2.3 Other sources.

Dermorphine is an opioid heptapeptide with the sequence Tyr-D-Ala-Phe-Gly-Tyr-Pro-Ser-NH_2 that was first extracted from the skin of Phyllomedusa sauvagei, a species of frog from South and Middle America[36,48]. Dermorphine is about one thousand times more potent than morphine in terms of activity. D-alanine present in this peptide is crucial for its activity.

Bombinins H are a group of peptides which are variants of bombinin, a peptide from the skin secretion of certain frogs[49]. Bombinin H4 has the sequence Ile-D-aile-Gly-Pro-Val-Leu-Gly-Leu-Val-Gly-Ser-Ala-Leu-Gly-Gly-Leu-Leu-Lys-Lys-Ile-NH_2[50] (Since isoleucine has two chiral centers, inversion of that at α-carbon atom results in D-alloisoleucine, represented as D-aile)[51].

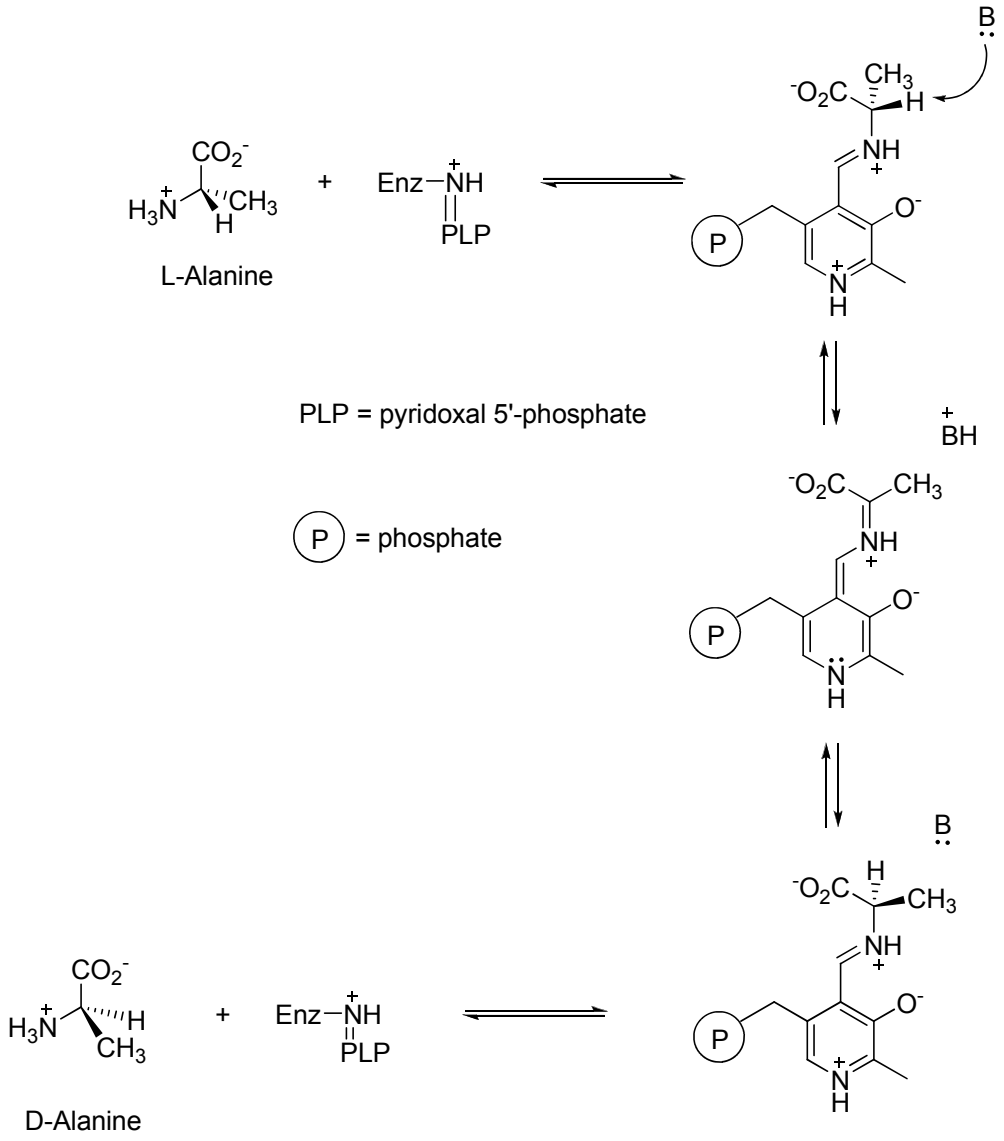
Two D-amino acid-containing neuropeptides have been isolated from the African giant snail *Achatina fulica*[36]. Achatin I, Gly-D-Phe-Ala-Asp, was isolated from its ganglia and atrium. Achatin I enhance activities of the snail's heart in two ways; first, this peptide increases impulse frequency and produces spike broadening of a heart excitatory neuron. In another way, this neuropeptide increases the amplitude and frequency of the heart beat. Achatin II (the corresponding peptide with an L-Phe) was also found but shows no biological activity. Another peptide is fullicin, Phe-D-Asn-Glu-Phe-Val-NH_2, which was also isolated from ganglia of the same snail and found to be a potent stimulator of the contraction of the penis retractor muscle of the snail.

3.2.4 Origins of D-amino acid-containing peptides.

D-amino acids can be synthesised by different mechanisms, depending on the organism where the D-amino acids are found.

Biosynthesis of peptidoglycan.

Alanine racemases are a group of pyridoxal 5'-phosphate (PLP)-containing enzymes, which use a pyridoxal cofactor to generate and stabilise an α-carbanion which, upon quenching, yields both L- and D-alanine (**Scheme 3.3**)[52].



Scheme 3.3: Racemization of alanine by alanine racemase

Other enzymes involved in peptidoglycan biosynthesis are diaminopimelate epimerase and glutamate racemase. Diaminopimelate transforms LL-diaminopimelate (LL-DAP) to *meso*-diaminopimelate (DL-DAP) (see **Figure 3.15** for DAP) while glutamate racemase generates D-Glu[53].

Products of other racemases.

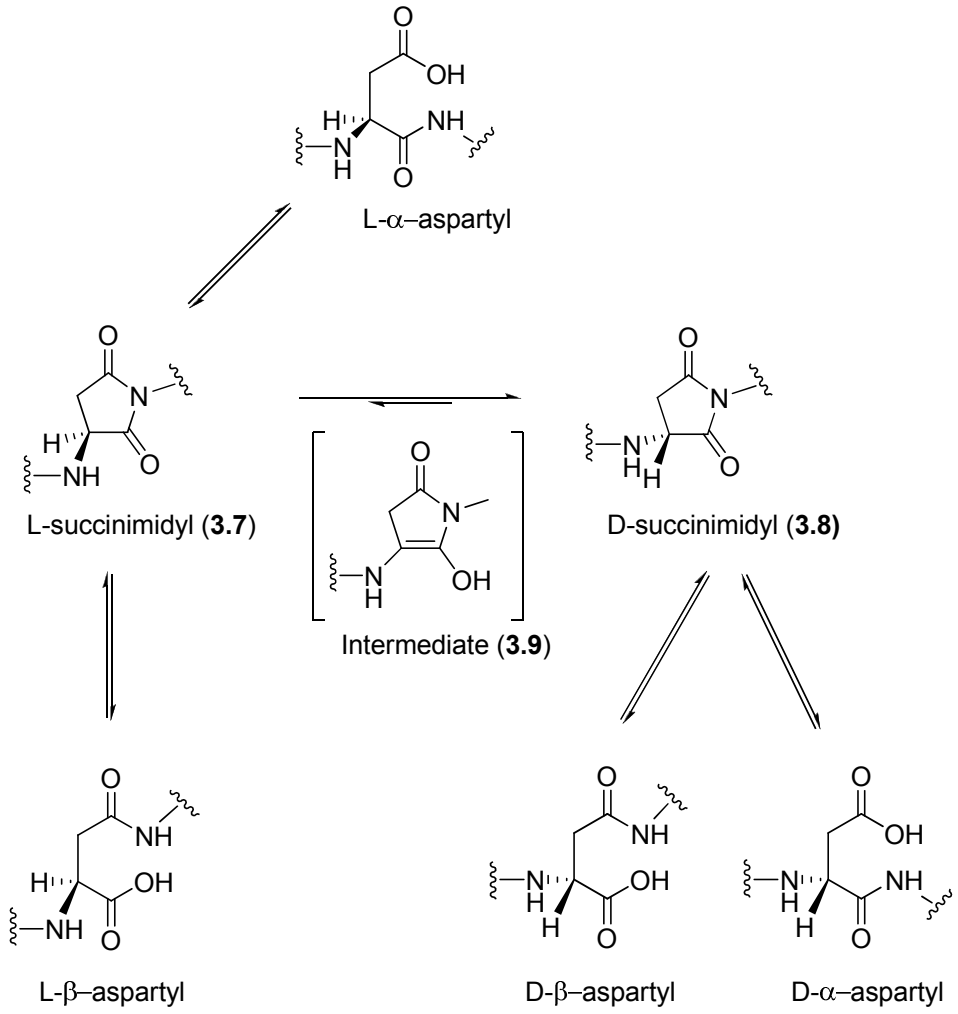
Besides alanine, racemases of other amino acids have been found. One of the most well studied is serine racemase, probably because D-serine is found in the brain of many vertebrates. Wolosker[54] purified serine racemase and studied its mechanism. Analysis showed that D-serine increased in direct proportion to L-serine decreasing, showing that L-serine was directly converted to D-serine by this enzyme, without a free intermediate (e.g. α–keto acid).

Chiral environment-inducing inversion of configuration.

α-Crystallin is a protein found mostly in eye lens, consisting of αA-crystallin and αB-crystallin subunits[55]. αA-crystallin is a peptide containing 173 amino acids. A D-amino acid is found in position 151 (Asp-151). In order to understand the stereo-inversion at Asp-151, the racemisation of Asp in a 12-amino acid peptide model corresponding to αA-crystallin (IQTGLDATHAER) was studied. The ratio of D-isomer to L-isomer (D/L ratio) of the Asp in a model peptide did not exceed 1.0 while in native protein, the D/L ratio at this position was found to be 5.7. This suggested that a chiral enviroment exists in the native higher order structure of human αA-crystallin which induces the inversion of L-Asp to D-Asp residues.

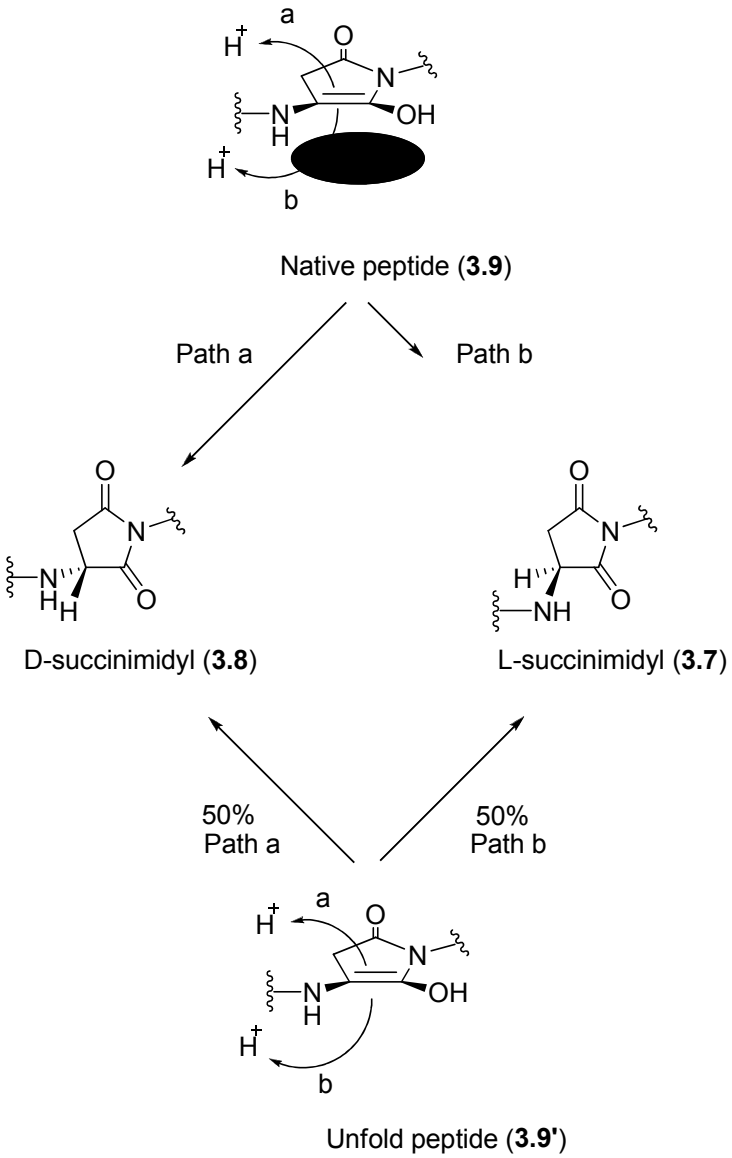
The proposed mechanism for inversion of configuration of Asp-151 is shown in **Scheme 3.4**.[55, 36] L-succinimidyl (**3.7**) can be converted into D-succinimidyl (**3.8**)

via achiral intermediate (**3.9**) by deprotonation and reprotonation. From either succinimidyl (**3.7** and **3.8**), hydrolysis can occur at either carbonyl group, resulting in 4 possible products.



Scheme 3.4: Formation of D-aspartyl unit in α-crystallin.

The key species in the predominance of the D-isomer of Asp-151 is intermediate **3.9**. The native higher order structure of the protein blocked the *Re*-face of **3.9**, resulting in protonation on the *Si*-face, giving the isomer **3.8** as a product (**Scheme 3.5**)[55].



Scheme 3.5: Explanation of unequal protonation leading to predomination of D-succinimidyl in α-crystallin.

3.3 Enzymes acting on D-amino acids and their applications.

Because of the chiral specificity of enzymes, most L-amino acid-containing peptides are inert to D-amino specific enzymes, and *vice versa*, very few enzymes work on D-amino acids and their derivatives.

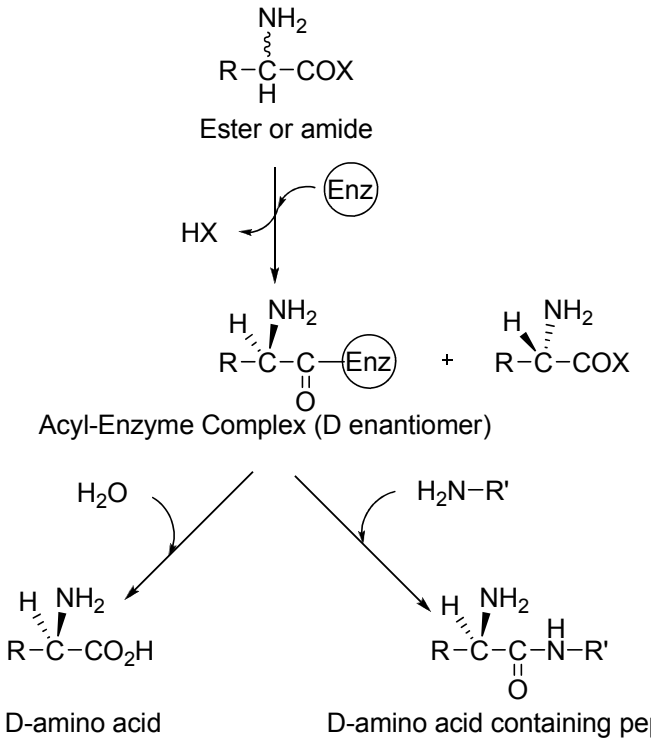
D-aminoacylases.

D-aminoacylases are a group of enzymes that specifically deacetylate *N*-acetyl-D-amino acids while they are not active against the corresponding L-amino acid derivatives[56] These enzymes are useful in the resolution of racemic *N*-acetyl amino acid.

D-aminopeptidases.

Aminopeptidases are enzymes which catalyze the stepwise cleavage of a single amino acid from the amino terminus of a peptide. Aminopeptidases acting on D-amino acid-containing peptides have been found in bacteria. D-aminopeptidases release a D-amino acid residue from the amino terminus of the peptide regardless of the stereochemistry of the second amino terminal amino acid[38].

D-aminopeptidases can be used in the synthesis of D-amino acids or D-amino acid containing peptides from racemic amides or esters. D-aminopeptidases can selectively form an acyl-enzyme complex with the D-enantiomer. Hydrolysis of this complex leads to enantiomerically pure amino acid. Alternatively, if this acyl-enzyme complex is reacted with a peptide, an *N*-terminal D-amino acid-containing peptide is formed (**Scheme 3.6**)[38] D-alaninamide amidase is an example of an enzyme from this group. It specifically hydrolyse D-alanineamide to D-alanine[57].



Scheme 3.6: Production of D-amino acids and D-amino acid containing peptides by D-aminopeptidase.

D-aminopeptidases have been applied in drug delivery. One strategy to deliver anticancer drugs to desired locations, and nowhere else, is to use an inactive prodrug that will be activated by enzymes located on targeted positions. Methotrexate (MTX) (**Figure 3.16 A**) is a drug widely used in chemotherapy. Different 2-α-aminoacyl derivatives of this molecule have been prepared and their suitability as prodrugs evaluated. D-Ala-MTX (**Figure 3.16 B**) was stable in serum but activated by D-aminopeptidase. This prodrug would thus be useful for application in antibody-directed enzyme prodrug therapy (ADEPT), where D-aminopeptidase is localized at the tumor site[58].

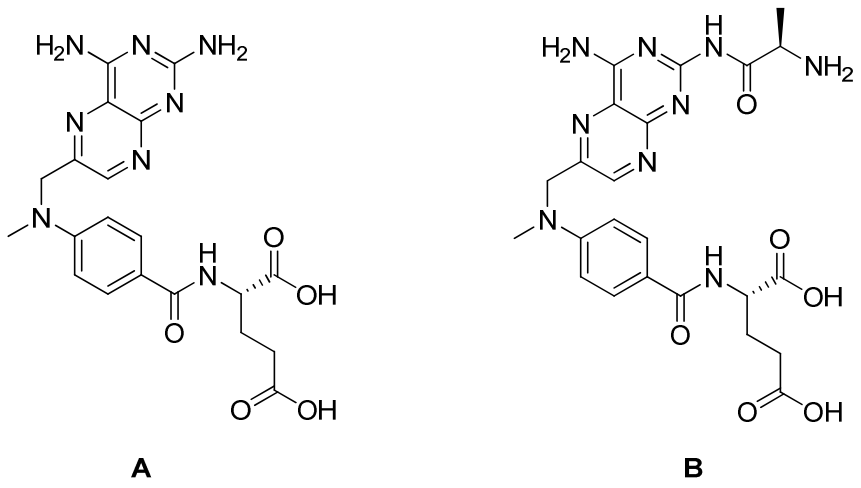


Figure 3.16: **(A)** Methotrexate; **(B)** 2-D-Ala- Methotrexate.

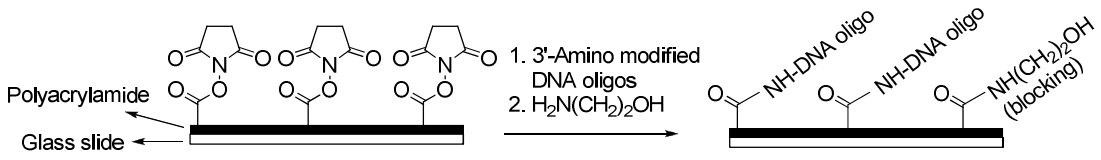
3.4 Result and Discussion.

3.4.1 DNA microarrays.

Microarrays used in this thesis were either prepared by printing amino modified DNA sequences onto CodeLink™ slides (for library **L 2.1, Figure 2.10**), or custom designed DNA microarrays (Agilent, for library **L 2.2 – L 2.4, Figure 2.11-2.13**). This section will describe only the DNA microarray prepared on CodeLink™ slides.

3'-Amino modified DNA oligos (18 bases long) with 12 bases complimentary to the PNA encoding oligos (6 bases were used as a spacer) were used for microarray fabrication. From the encoding strategy presented in **2.4.2** (**Table 2.2**), library **L 2.1** had 125 different PNA sequences therefore 125 complementary DNA sequences were used for decoding. The success of the printing process was monitored by printing of a 3'-amino modified FAM labelled DNA sequence (so called printing reference). Each DNA sequence was printed with up to 7 replicates per slide, except the printing reference which was printed 21 times. As the result, each slide contained 896 spots [(7x125)+21].

The microarray fabrication was performed using a Robot Microarrayer (Genetix QMini, UK). The process is summarised in **Scheme 3.7**. The surface of a CodeLink™ slide was modified by printing on a 3-D polyacrylamide gel matrix, containing *N*-hydroxysuccinimide (NHS) active esters. 3'-Amino modified DNA oligos were immobilized on this slide upon printing, with unreacted NHS active esters being blocked by reacting with ethanolamine.



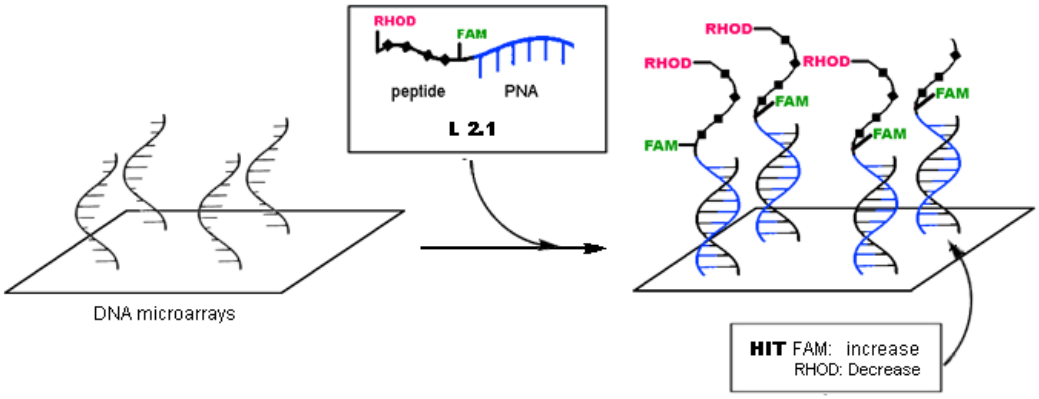
Scheme 3.7: Immobilization of DNA oligos onto CodeLink™ slides.

3.4.2 PNA-encoded D-amino acid-containing peptide libraries as potential substrates for chymopapain.

PNA encoded D-amino acid-containing peptides libraries (**L 2.1** and **L 2.2,** Chapter 2) were generated and treated with chymopapain in order to identify possible D-amino acid containing peptides as substrates of this enzyme.

3.4.2.1 Screening of library L 2.1.

The analysis of library **L 2.1** is summarised in **Scheme 3.8**. After incubation of the "D" library with chymopapain, cleaved members would have rhodamine B removed. As a result, the FRET would be interrupted and therefore an increase in FAM (FRET donor) intensity would result. As explained earlier, FAM intensity must be considered in comparison with the control, which was rhodamine B in this case. Therefore, the sequences with higher ratios of FAM (treated library) to rhodamine B (control library) were of interest.



Scheme 3.8: Screening of a FRET D-amino acid containing peptide library (**L 2.1**).

One portion of library **L 2.1** (where AA^4 was L-Ala) was screened as a potential substrate for chymopapain. From the 125 library members, the 12 sequences with the highest FAM/RHOD ratios were identified (**Figure 3.17**) (n = 2).

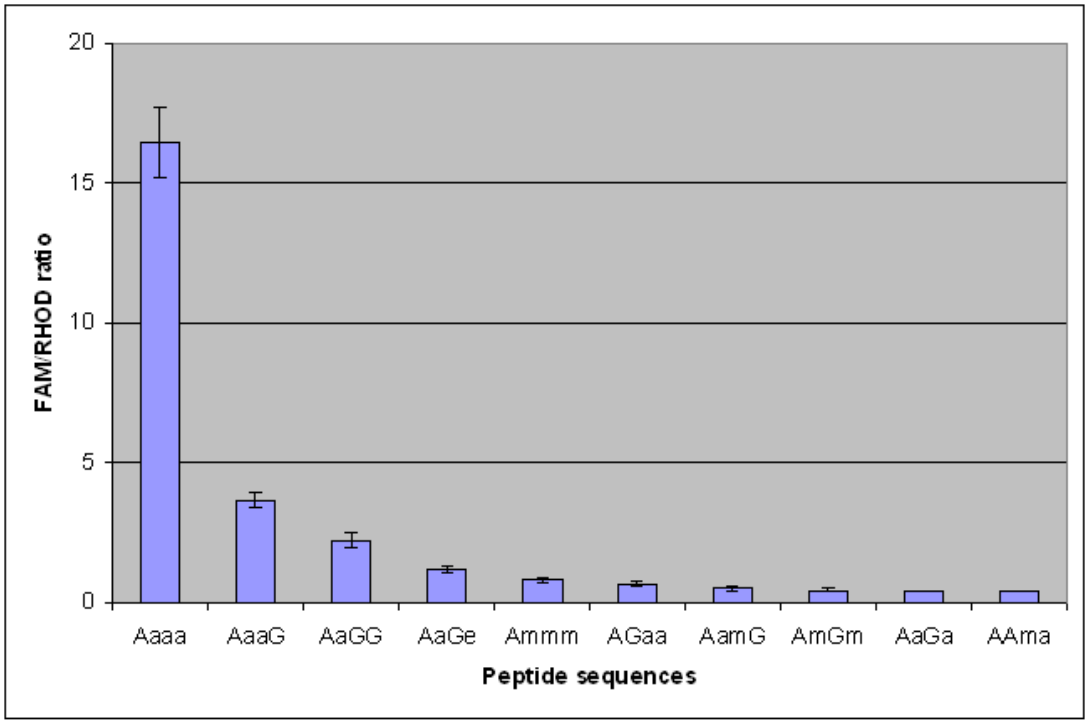


Figure 3.17: Cleaved peptide sequences from **L 2.1**.

From **Figure 3.17**, it can be concluded that, of the 125 members screened, three sequences, namely L-Ala-D-Ala-D-Ala-D-Ala, L-Ala-D-Ala-Gly-Gly and L-Ala-D-Ala-D-Ala-Gly, showed significant increases in the FAM/RHOD ratio. These

sequences were considered positive "hits" and one of these sequences was selected for further study.

The FRET peptide from the "hit" peptides, Methyl Red-Ahx-L-Ala-D-Ala-D-Ala-D-Ala-D-Lys(FAM)-NH_2 (peptide **3.10**), was synthesised. The FRET-pair were identical to those described in section 3.1.6. In peptide **3.10**, fluorescein was conjugated to the tetrapeptide *via* D-Lys, the same linker used in library **L 2.1** (**Figure 3.18**).

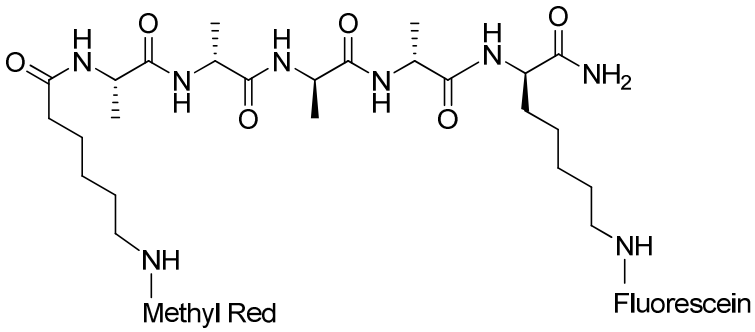


Figure 3.18: Peptide **3.10** (Methyl Red-Ahx-L-Ala-D-Ala-D-Ala- D-Ala-D-Lys(Fluorescein)-NH_2), a FRET peptide used in the enzymatic evaluation of chymopapain against D-amino acid containing peptide.

Peptide **3.10** was subjected to enzymatic digestion with chymopapain at 37°C and pH 7.4 and the cleavage of the peptide was followed as in section **3.1.6**. The fluorescence profile of the hydrolysis is shown in Figure **3.19**. No increase in fluorescent intensity was detected. The fluorescence profiles suggested that peptide **3.10** was not digested by chymopapain.

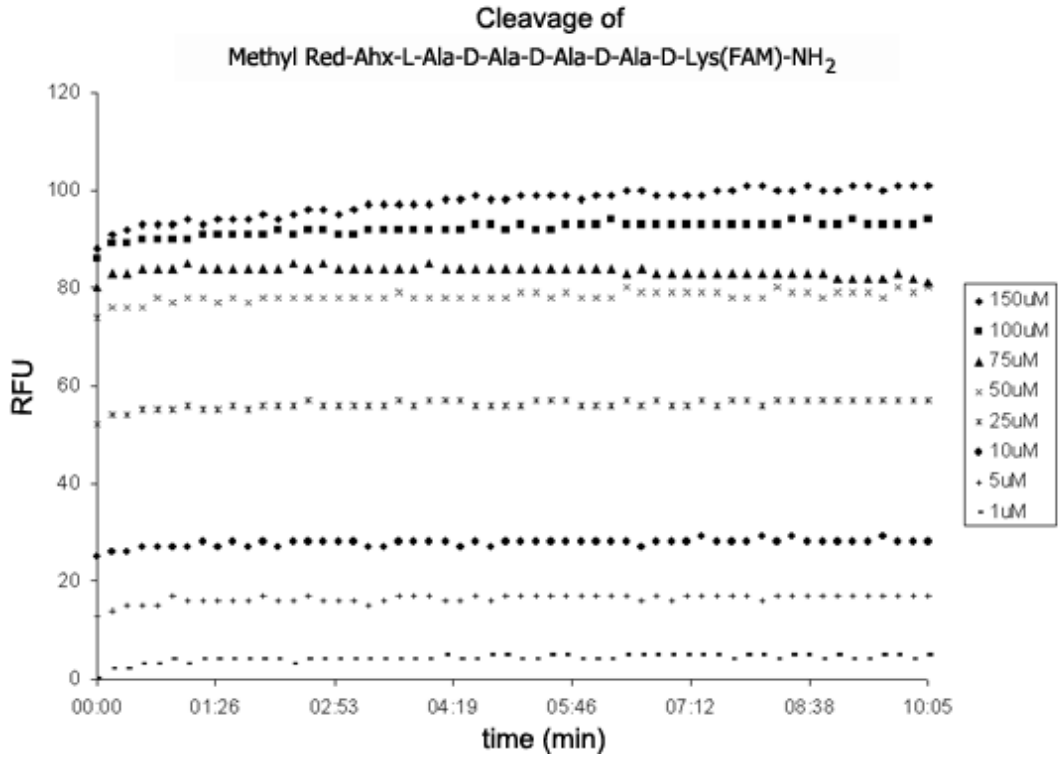


Figure 3.19: Fluorescence profiles (relative fluorescence units [RFU] vs. time [min]) following the hydrolysis of peptide **3.10** by chymopapain at pH 7.4.

In order to confirm this result, peptide **3.10** (100 µM) was incubated with chymopapain (10 µM) at 37°C and pH 7.4 and the cleavage of peptide **3.10** was followed by analytical HPLC analysis at 515 nm at different reaction times (0 h, 1.5h and 18 h) (**Figure 3.20**). The HPLC analysis showed only the peak of the starting material, confirming that peptide **3.10** was not digested by chymopapain.

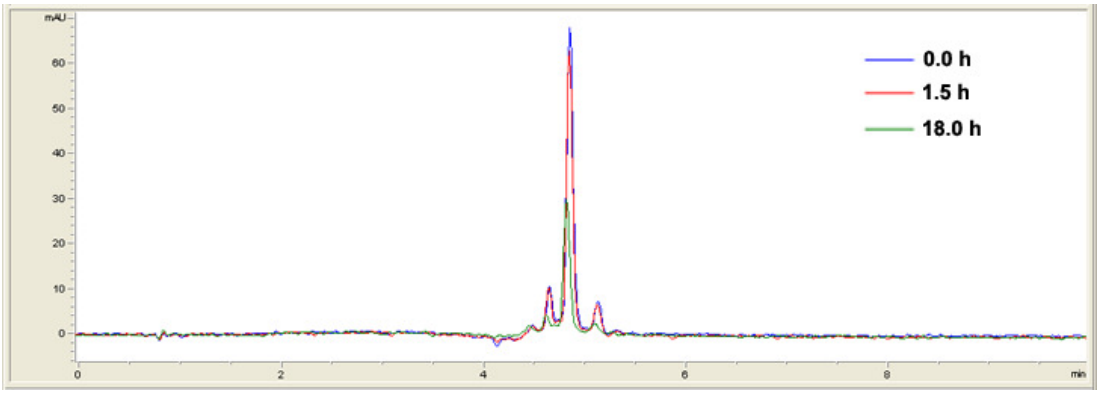
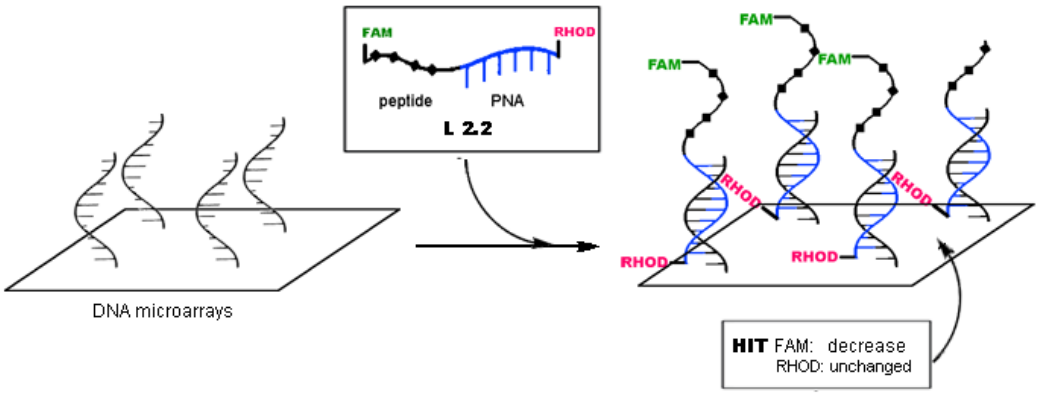


Figure 3.20: HPLC traces recorded at 515 nm of peptide **3.10** following treatment with 5 µM chymopapain at 37°C and pH 7.4 at different reaction time.

Thus although the tetrapeptide L-Ala-D-Ala-D-Ala-D-Ala showed the highest FAM/RHOD ratio (**Figure 3.17**), the cleavage of this peptide could not be detected when it was used as a single substrate. The possible explanation for this result is that, because methyl red was used as a dark quencher in peptide **3.10** while rhodamine B was used in the library, it is possible that methyl red could be influencing the cleavage of the peptide. Therefore the cleavage of "hit" sequence (peptide **3.10**) could not be observed.

3.4.2.2 Screening of library L 2.2.

Library **L 2.2** was a 1296 membered PNA-encoded tetrapeptide library which contained FAM on the amino terminus of the peptide and rhodamine B amide on the amino terminus of the PNA chain. Therefore cleaved sequences of **L 2.2** members would result in removal of FAM. The "hit" sequences were then identified by a decrease in the ratio of FAM to control chromophore (rhodamine B) (**Scheme 3.9**).



Scheme 3.9: Screening of a D-amino acid containing peptide library (**L 2.2**) as chymopapain substrate.

Figure 3.21 (a) shows the FAM/RHOD ratio from library **L 2.2** before ("control") and after treatment with chymopapain. Both experiments did not show

any difference in FAM/RHOD ratios. As presented in **Figure 3.21 (b)**, the difference of FAM/RHOD ratios of the control library and FAM/RHOD ratios of chymopapain treated library were mostly less than 0.1. In addition, as presented in **Figure 3.21 (a)**, peptides containing D-Pro-D-Lys at positions AA^4-AA^3 or AA^3-AA^2 (shown in the boxes) had a low FAM/RHOD ratio compared to the rest of the libraries (presumably due to the conformation of the peptide and the beta-turn induced by the proline residue).

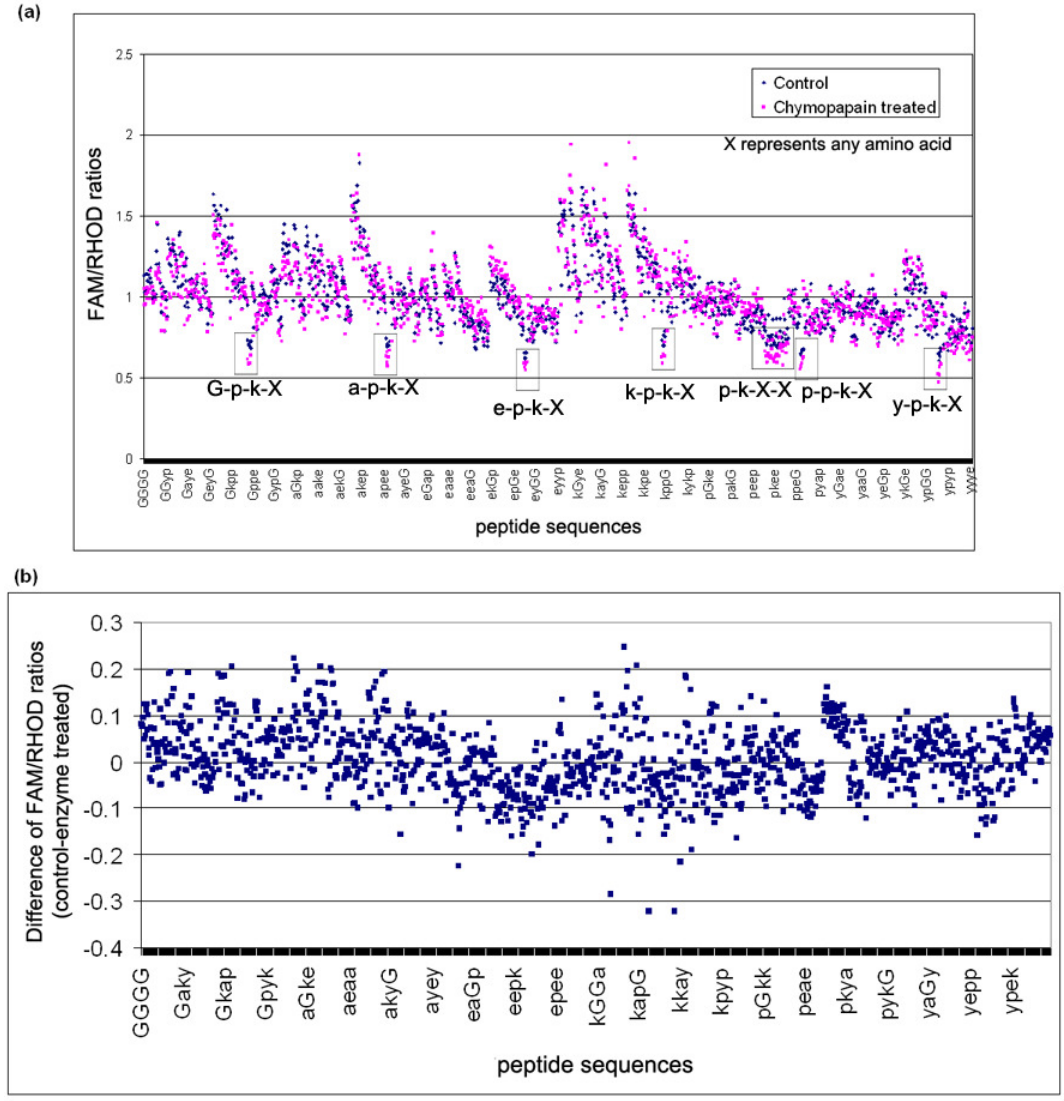


Figure 3.21: (a) FAM/RHOD ratio from library **L 2.2** before (blue) and after treatment with chymopapain (pink) **(b)** The difference of FAM/RHOD ratios of control library and FAM/RHOD ratios of chymopapain treated library.

These data are also presented in another format, in which the FAM intensities obtained from the library after the treatment with chymopapain are plotted against FAM intensities obtained from the library before the treatment with chymopapain (so called "control") (**Figure 3.22**). A group of sequences showed higher FAM intensities following chymopapain treatment, opposite to that expected from the cleaved sequences (presented in the box in **Figure 3.22**). Members of this group contained D-Pro-D-Lys at positions AA^4-AA^3 or AA^3-AA^2. However, when the FAM/RHOD ratios were used instead of FAM intensities, the data plotted in diagonal lines were obtained meaning that the treatment with chymopapain did not change the FAM/RHOD ratios of the library members (**Figure 3.23**). This result underlines the importance of an internal control (rhodamine B in this case) in a library containing large number of members.

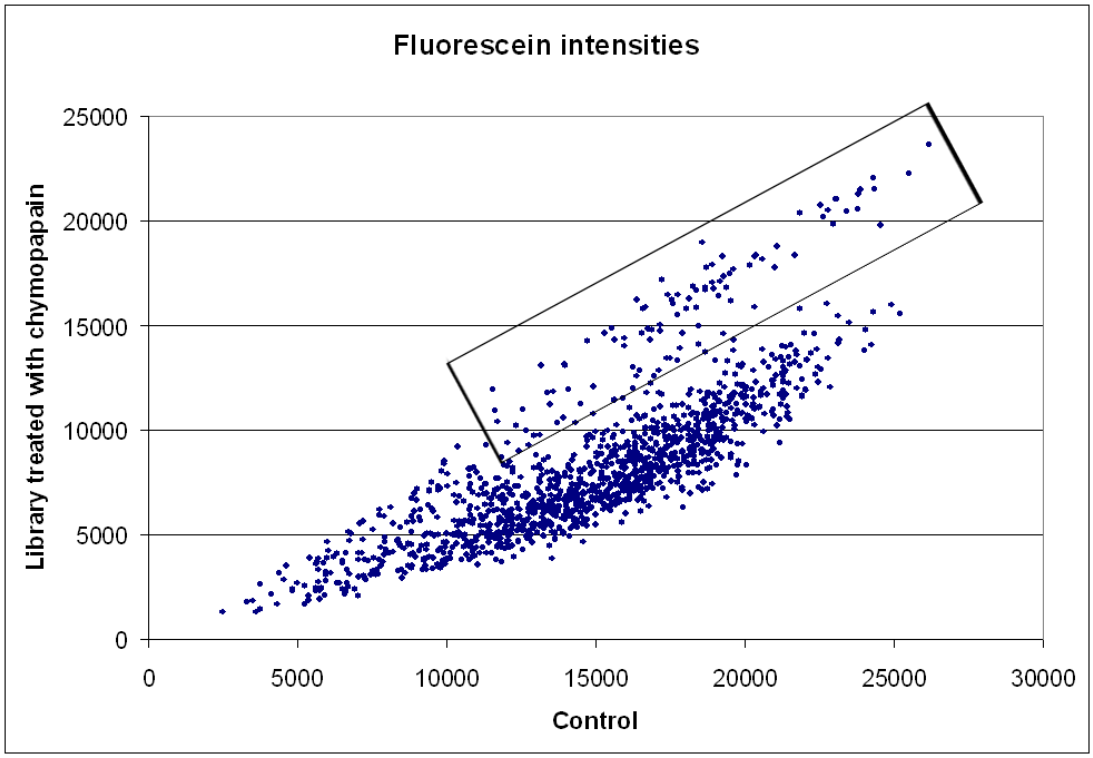


Figure 3.22: Fluorescein intensities obtained from library **L 2.2**, comparing chymopapain treated library (y-axis) against control (x-axis).

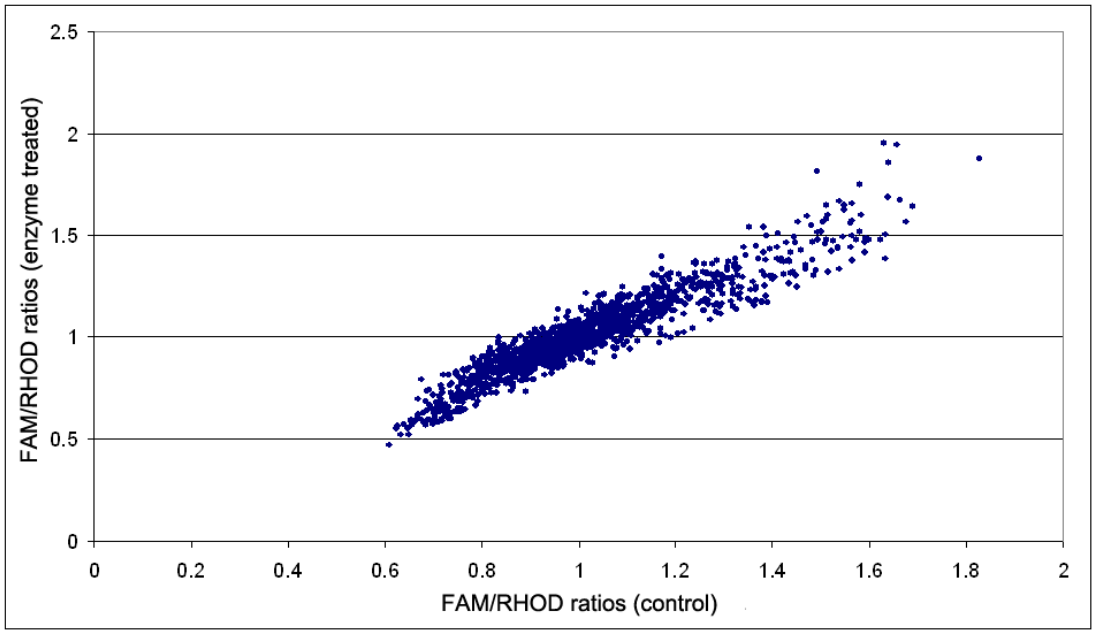


Figure 3.23: FAM/RHOD ratios obtained from library **L 2.2**, after treatment with chymopapain (y-axis) against control library (x-axis).

FRET intensities from library **L 2.2**, treated with chymopapain, are shown in **Figure 3.24**. FRET intensities were obtained from excitation of the microarrays at 488 nm (FAM excitation wavelength) and the emission intensities measured at 575 nm (RHOD emission wavelength). If FRET takes place in library **L 2.2**, the cleaved peptide would remove the donor chromophore FAM, hence a decrease in emission intensity because of FRET would be expected. From **Figure 3.24 (a)**, the result showed that no decrease in FRET intensity was observed when the library was treated with chymopapain. **Figure 3.24 (b)** also showed that, while FRET intensities from most of the library members did not change, some members had FRET intensities that increased after the treatment with chymopapain.

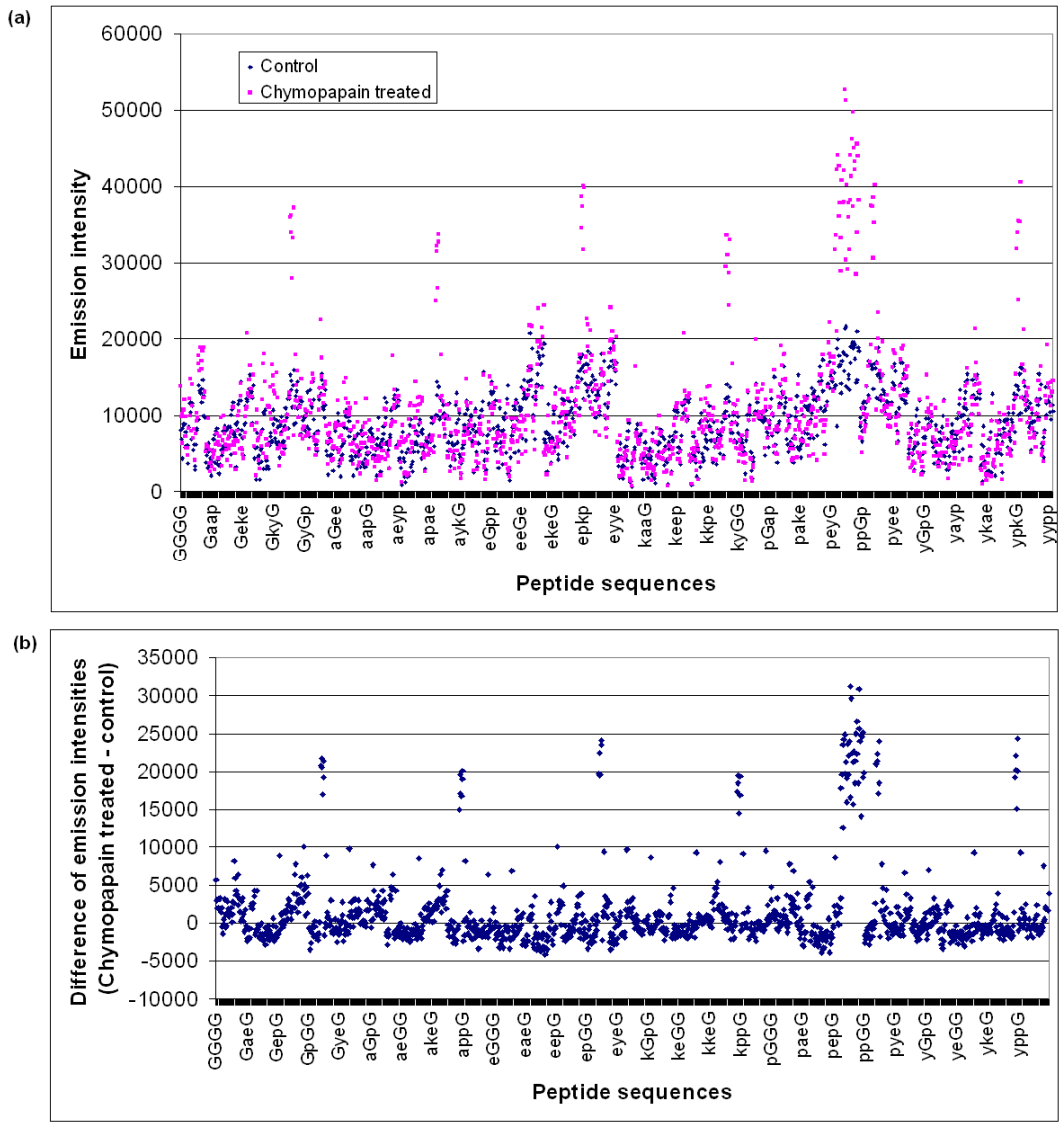


Figure 3.24: **(a)** FRET intensity obtained from library **L 2.2** before (blue) and after (pink) treatment with chymopapain **(b)** The difference of FRET intensities obtained from **L 2.2**.

It was found that many library members containing D-Pro-D-Lys at the positions AA^4-AA^3 or AA^3-AA^2 responded differently to the treatment with chymopapain compared to the rest of the library. However, the responses from those peptides were apparently not due to the cleavage by the enzyme because the intensities corresponding to both chromophores were always found to be increased after the treatment of the enzyme. The reason for this behaviour is unclear. One possible explanation is the 3-dimensional arrangement of the peptide sequences. Having D-Pro as one of amino acids on the specific locations of those sequences, the

sequences possibly folded back in a beta-turn stabilized structure, therefore they had a different 3-dimensional arrangement compared to the rest of the library and responded differently to microarray scanning.

Since peptide sequences containing D-Pro-D-Lys at positions AA^4-AA^3 or AA^3-AA^2 behaved differently compared to the rest of the library, this sequence was selected for further studies. A FRET tetrapeptide FAM-Ahx-D-Ala-D-Pro-D-Lys-Gly-D-Lys(Methyl Red)-NH_2 (peptide **3.11**, **Figure 3.25**) was synthesised for enzymatic evaluation.

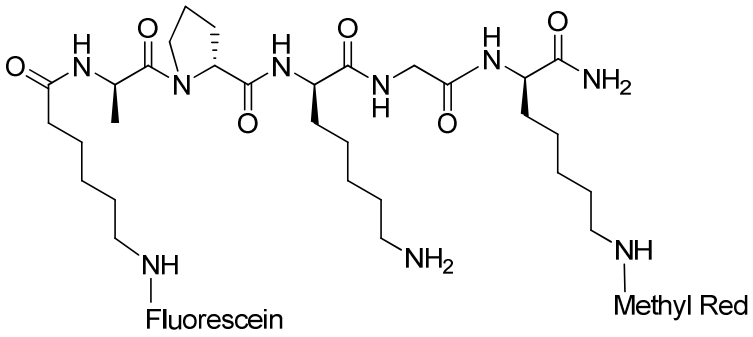


Figure 3.25:Peptide **3.11**, (Fluorescein-Ahx- D-Ala-D-Pro-D-Lys-Gly-D-Lys(Methyl Red)-NH_2), a FRET peptides used in enzymatic evaluation of chymopapain.

Peptide **3.11** (100 μM) was subjected to enzymatic digestion with 5 μM chymopapain at 37°C and pH 7.4. After 24 h the reaction (and a control, in which 100 μM of Peptide **3.11** was incubated at 37°C and pH 7.4 without chymopapain) were analysed by analytical HPLC (**Figure 3.26**) and showed that peptide **3.11** was not cleaved by chymopapain.

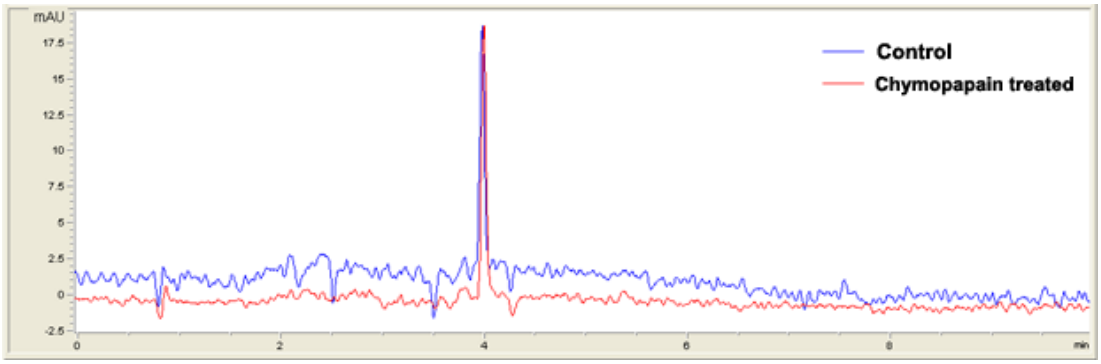


Figure 3.26: HPLC traces recorded at 515 nm of peptide **3.11** following treatment with 5 μM chymopapain at 37°C and pH 7.4 for 24 h. (red) and control (blue).

The result that peptide **3.11** was not digested by chymopapain thus discarded the possibility that dipeptide D-Pro-D-Lys is the target of chymopapain. Therefore no D-amino acid-containing peptide from library **L 2.2** could be identified as substrates of chymopapain.

3.4.3 PNA-encoded D-amino acid-containing peptide libraries as potential substrates for Gumby.

Gumby is a novel papain-like cysteine protease discovered by Dr. Sabine Cordes's research group. Substrate for this protease is still unknown therefore the protocol previously presented could be used for the discovery of specific substrate(s) for this protease.

Library **L 2.2** was screened as potential substrate for wild-type and mutant gumby. After being denatured by heating at 90°C for 20 min, library **L 2.2** (600 pmol/experiment) was treated with wild-type and mutant Gumby (11.4 pmol). The enzymatic reaction was allowed to proceed for 22 hours at 37°C. The reaction mixture was then hybridised onto DNA microarray and analysed in a similar fashion to that described in section 3.4.2.

According to **Scheme 3.9**, if proteolytic activity took place, the FAM/RHOD ratio should decrease compared to the control. This ratio, from library **L 2.2** treated with wild-type Gumby, showed that none of the library members had different ratios from the majority of the library (**Figure 3.27**) (similar to the control).

Mutant Gumby showed the same result to that of wild-type. FAM/RHOD ratios of all library members are similar to that of control (data not shown). Therefore it can be concluded that none of the PNA-encoded library **L 2.2** was cleaved by wild-type or mutant Gumby.

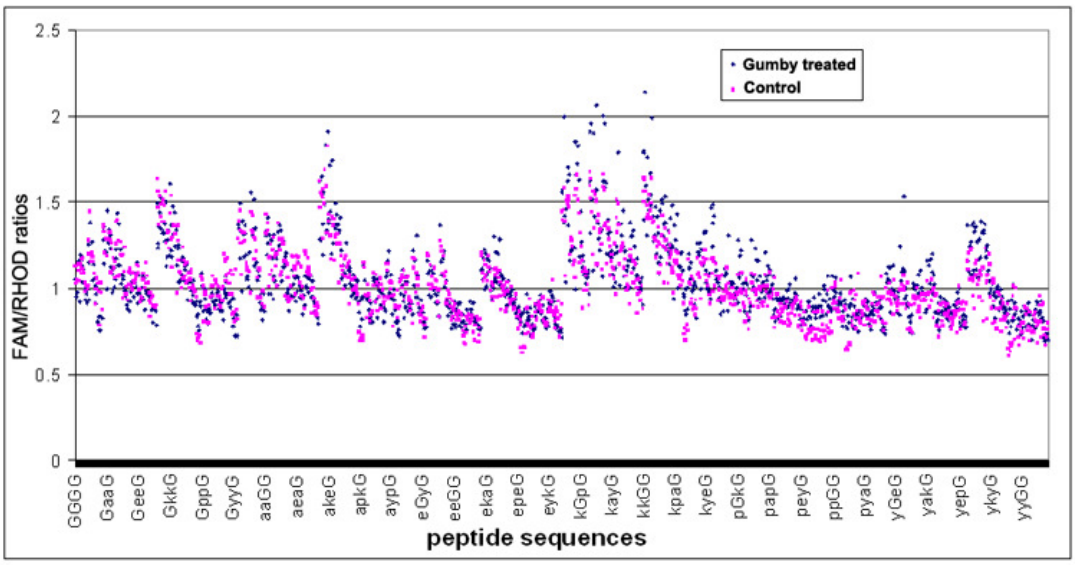


Figure 3.27: FAM/RHOD ratio obtained form library **L 2.2** after being treated with wild-type Gumby and control.

3.5 Conclusions.

The success of the PNA encoding strategy in the combinatorial libraries was demonstrated. With this strategy, specific substrates for chymopapain could be identified from a 10,000-membered FRET based hexapeptide library. The affinity and turnover number showed that these peptides are an ideal substrate for chymopapain. The same strategy was also used for libraries of D-amino acid-containing peptide for the discovery of new substrate(s) for a number of enzymes. D-

alanine might be the preferable amino acid of chymopapain, however, the digestion of a peptide containing D-alanine by chymopapain was not successfully demonstrated from any of the libraries evaluated..

A non-FRET PNA-encoded library of D-amino acid containing peptide was screened for substrates of chymopapain and a novel protease Gumby. However, no library member was found to be a potential substrate for any protease used.

Chapter 4: High-throughput screening of an encoded library of peptides as potential substrates for "self-promoting" Hüisgen 1,3-dipolar cycloadditions.

4.1 Introduction.

4.1.1 Hüisgen 1,3–dipolar cycloadditions.

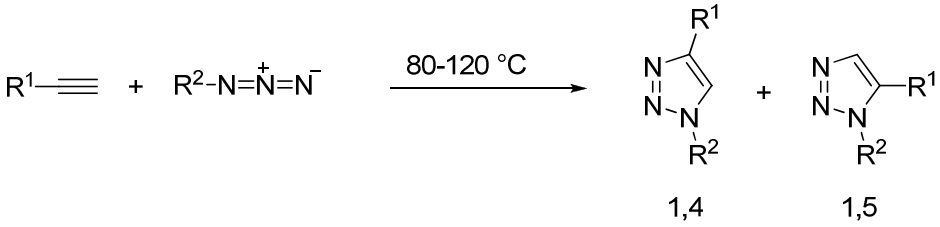
In 2001, the term "Click Chemistry" was introduced by Sharpless[59]. It refers to a series of organic reactions that display certain criteria; fast, stereospecific, reactions that take place under mild reaction conditions, giving product(s) in high yield with inoffensive by-products and with product isolation by simple nonchromatographic methods. "Click reactions" include many classes of transformation; cycloaddition, nuclephilic substitution, carbonyl chemistry of the "non-aldol" type and additions to carbon-carbon multiple bonds.

Carbon-carbon unsaturated bonds

Carbon-carbon unsaturated bonds are relatively more reactive than carbon-carbon single bonds making olefins and carbon-carbon triple bonds attractive starting materials to the synthetic chemist. Because of their reactivities, compounds containing carbon-carbon unsaturated bonds are good reagents for "Click" transformations.

Among the number of reactions considered, the Hüisgen 1,3–dipolar cycloaddition is one of the most well known due to its extensive applications in various fields such as chemistry, biology and materials science. This cycloaddition involves 1,3-dipolar compounds and diatomic dipolarphiles, giving rise to 5-

membered heterocycles. 1,3-Dipolar molecules can be compounds such as azides, nitrile oxides, and ozone, while alkynes and nitriles are examples of dipolarphiles. One of the most studied couples may be the azido and alkyne derivatives, giving 1,2,3-triazoles as products. Although it is highly exergonic ($\Delta G^{\circ} \approx$ -255 kJ/mol), a high temperature is necessary to initiate the reaction ($\Delta G^{\ddagger} \approx$ +108 kJ/mol)[60], and a mixture of both 1,4- and 1,5- disubstitue triazoles is obtained (**Scheme 4.1**). However, in the presence of copper(I), the cyclization of terminal alkyne develops much faster and in a regiospecific manner, giving only 4-substituted 1,2,3-triazole as a product[61].

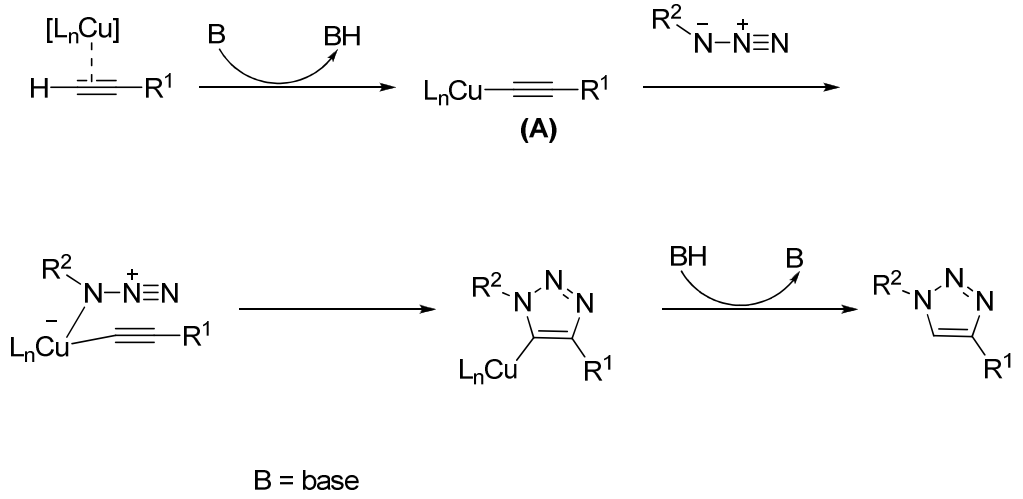


Scheme 4.1: Non-regiospecific Hüisgen dipolar cycloaddition.

Beside copper(I), this reaction can also be catalysed by ruthenium(II), $Cp^{*}RuCl(PPh_3)_2$ for example[62], although the ruthenium(II) catalysed reaction gives 5-substituted triazoles.

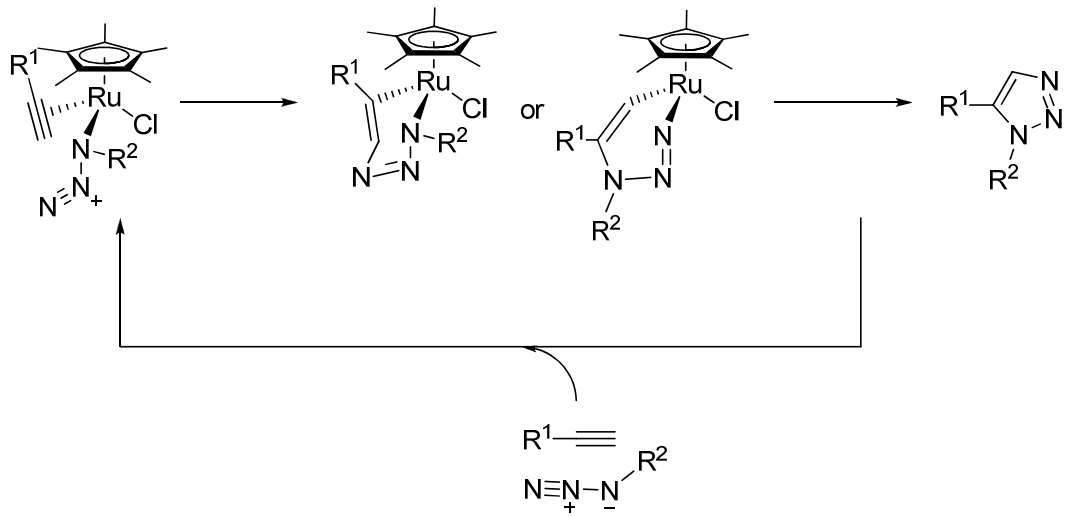
4.1.2 Mechanism of metal-catalyzed Hüisgen Cyclizations.

The mechanism of the copper(I) catalyzed Hüisgen Cyclization in case of terminal alkynes is represented in **Scheme 4.2**. Alkyne firstly forms a π complex with copper(I), resulting in a lowering of its pK_a. After deprotonation, copper(I) acetylide **(A)** regioselectively coordinates to its azide counterpart. Intramolecular cyclization gives the copper(I)-coordinated 1,2,3-triazole, which is then converted to the 1,2,3-triazole by reprotonation.



Scheme 4.2: Mechanism of Cu(I)-catalyzed Hüisgen dipolar cycloaddition[61].

The ruthenium(II)-catalyzed Hüisgen cyclization is different from the copper(I) catalyzed one. Thus 1,4- or 1,5-disubstitued triazole can be obtained selectively depending upon the ruthenium(II) complex used. For example, in the cyclisation of benzyl azide and phenylacetylene, $Ru(OAc)_2(PPh_3)_2$ give the 1,4-disubstituted triazole while $Cp^*RuCl(PPh_3)_2$ and $Cp^*RuCl(NBD)$ gave the 1,5-disubstituted cycloadduct[62]. Moreover, with ruthenium(II) catalysts, internal alkynes could be used while the copper(I)-catalyzed only works with terminal alkynes. The proposed catalytic cycle of ruthenium(II) leading to the 1,5-disubstituted product is represent in **Scheme 4.3**.



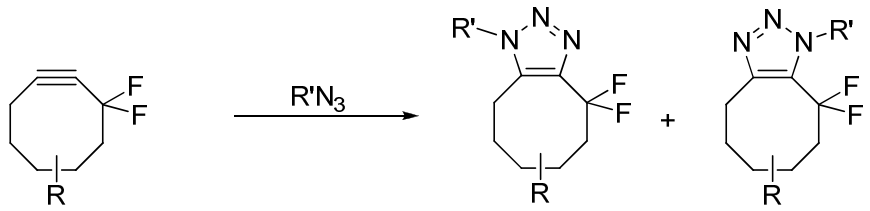
Scheme 4.3: Mechanism of Ru(II)-catalyzed Hüisgen dipolar cycloaddition[62].

4.1.3 Catalyst-free Hüisgen cycloadditions.

Hüisgen cycloaddition has been applied in a number of studies. However, one limitation is that it requires metal catalysis. Since these catalysts are often toxic to living systems, many reaction strategies have been proposed in order to achieve this cycloaddition process under catalyst-free conditions.

4.1.3.1 Difluorinated Cyclooctynes.

Difluorinated cyclooctyne have been considered as an alternative alkyne for Hüisgen cycloadditions. Cyclooctyne is the smallest stable cycloalkyne and it is strained to the order of 75 kJ/mol (i.e. this is "gained" when it reacts). As a result, the activation energy needed to achieve cycloaddition is very low[63]. Presentation of the two fluorines at the propargylic position makes the alkyne more electron deficient, increasing interaction with the electron-rich azide counterpart. By combining the two phenomena, ring strain and electron deficiency, difluorinated cyclooctyne can be coupled with azides at ambient temperature without the need for any metal catalyst (**Scheme 4.4**)[60].



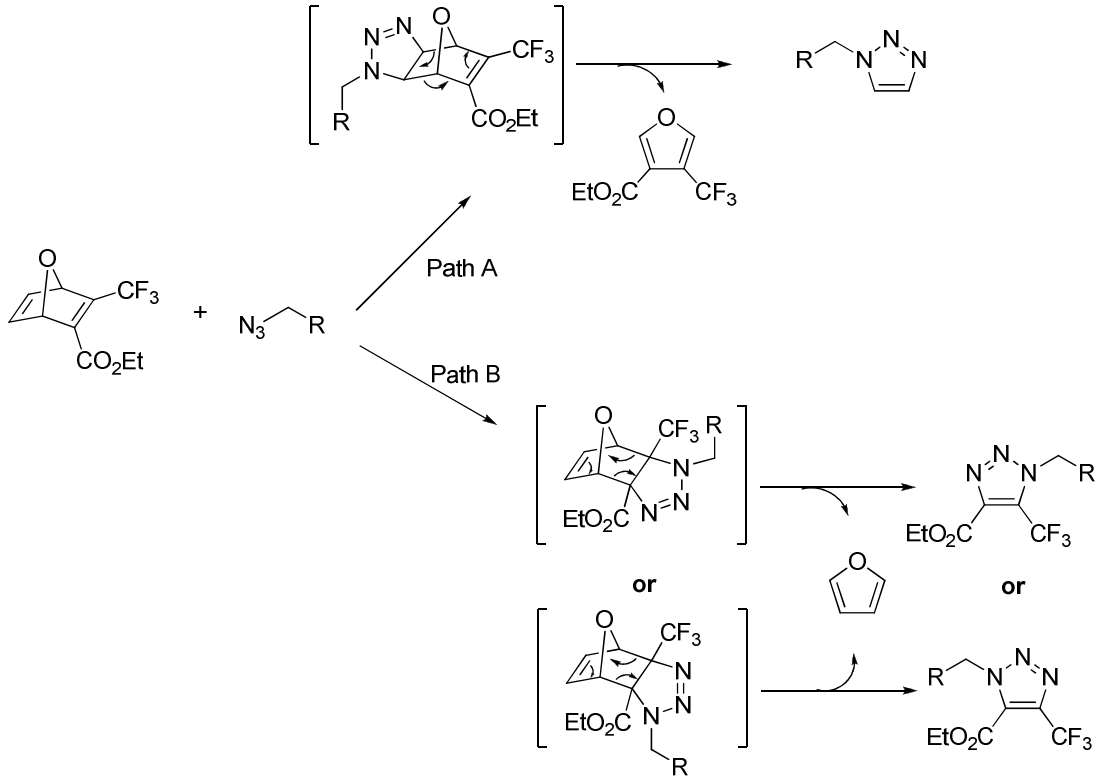
Scheme 4.4: Formation of 1,2,3-triazole from difluorinated cyclooctyne and azide[60].

Due to the possibility of doing the "Click reaction" without any transition metal catalyst, this cycloalkyne could be used in cells. Three biotinylated derivatives of difluorinated cyclooctyne were incubated with cells bearing azides on their surfaces. After staining with FITC-labeled avidin, cells could be visualized, with the cyclization taking place within 60 minutes[60].

Although difluorinated cyclooctyne derivatives are efficient reagents for Hüisgen cycloaddition, their preparation is difficult and the desired compounds are only obtained in low yields.

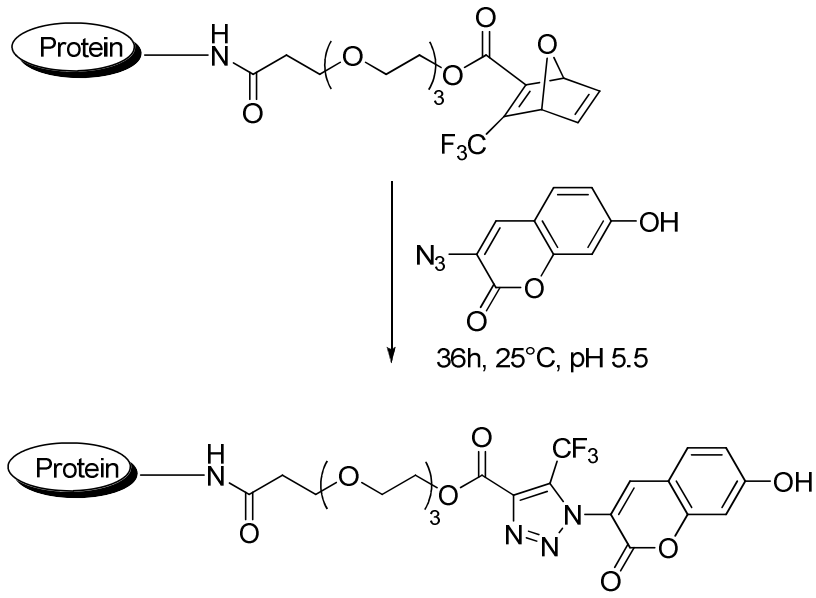
4.1.3.2 Oxanorbornadiene[64].

As a product of the Diels-Alder reaction between a furan and an alkyne, oxanorbornadiene also has a relatively strained ring. With the presentation of electron withdrawing substitutions, this bicyclic compound can also undergo Hüisgen cycloaddition with azides without transition metal catalysts, leading to trisubsitiued 1,2,3-triazoles. The mechanism represented in **Scheme 4.5** shows that tandem [3+2] cycloaddition-retro-Diel-Alder can take place at either alkene on oxanorbornadiene. However, cyclization *via* the more electron deficient (path B) is more preferable, therefore, 1,4,5- trisubstituted 1,2,3-triazoles are the major products.



Scheme 4.5: Reaction pathways for the formation of 1,2,3-triazoles from oxanorbornadiene and azide[64].

This strategy was applied in biological system[64]. A model protein was functionalized with oxanorbornadiene, and then reacted with 3-azido-7-hydroxycoumarin (without a metal catalyst) resulting in a triazole-conjugated coumarin, which could be visualized (**Scheme 4.6**).



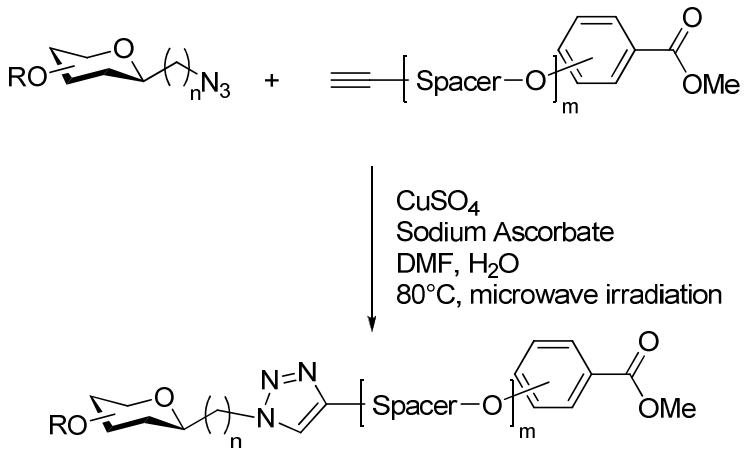
Scheme 4.6: Application of the oxanorbornadiene to catalyst-free Hüisgen cycloaddition in biological systems.

4.1.4 Applications of Hüisgen cycloadditions.

Due to its efficiency and user-friendliness, copper-catalysed Hüisgen cycloadditions have been used in various fields of chemistry from drug discovery, to material science. Some examples of these applications are discussed below.

4.1.4.1 Multivalent carbohydrates and peptides.

Multivalent sugars are of synthetic interest because they can interact with protein partners much more strongly than monovalent ones. Dendrimeric sugars have been synthesised using a copper-catalyzed Hüisgen cycloaddition between alkyne dendrimers and azido sugars, as summarized in **Scheme 4.7**[65]. It also demonstrates that alkynes and azides functional groups could be swapped between the two coupling partners.

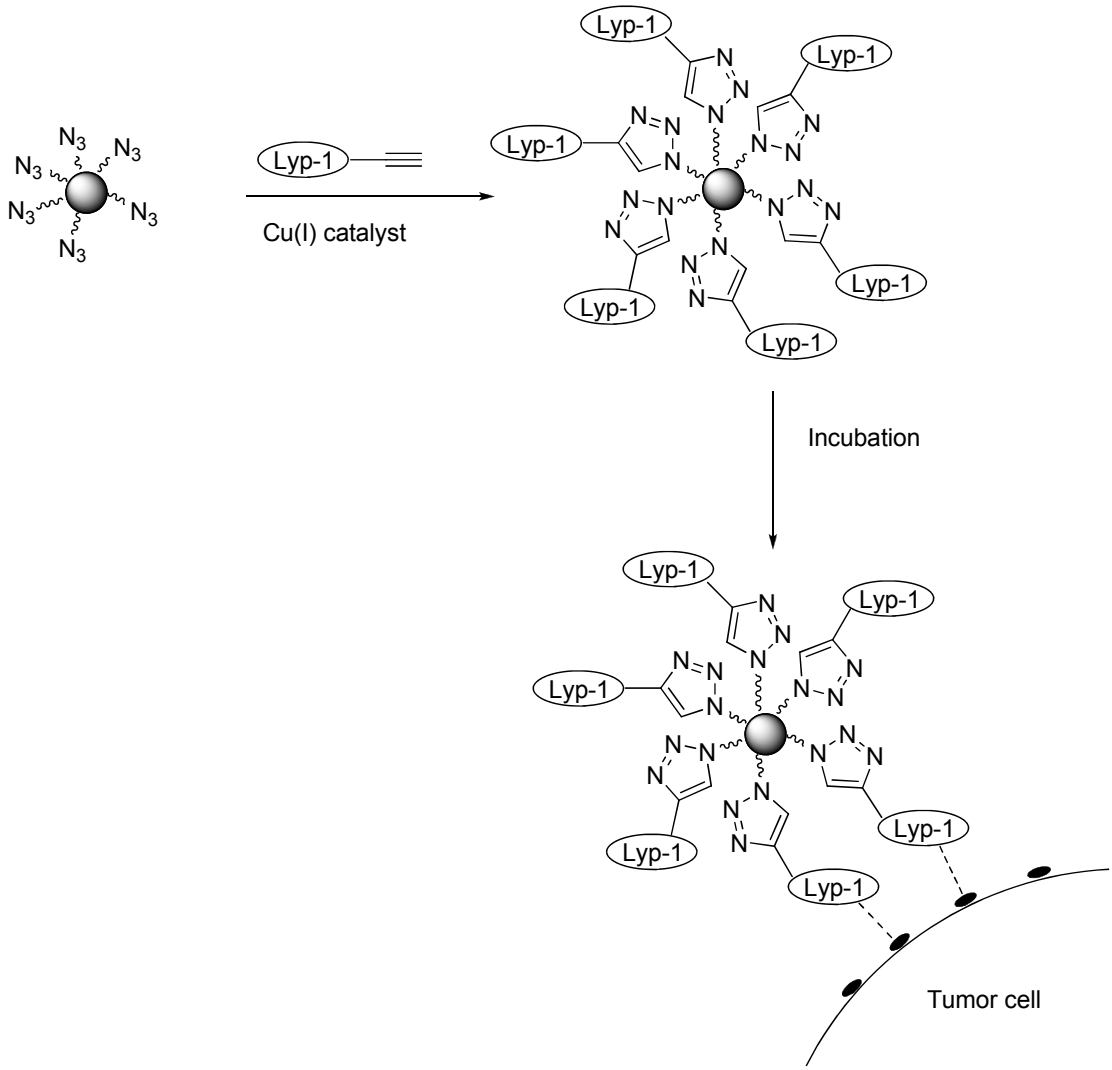


Scheme 4.7: Synthesis of multivalent sugars by Hüisgen cycloaddition.

In **Scheme 4.7**, with the same dendrimeric alkyne, azido sugars could be replaced with azido peptides, yielding multivalent dendrimeric peptides, that show enhanced biological properties compared to the original peptides[65].

4.1.4.2 Cellular targeting nanoparticles.

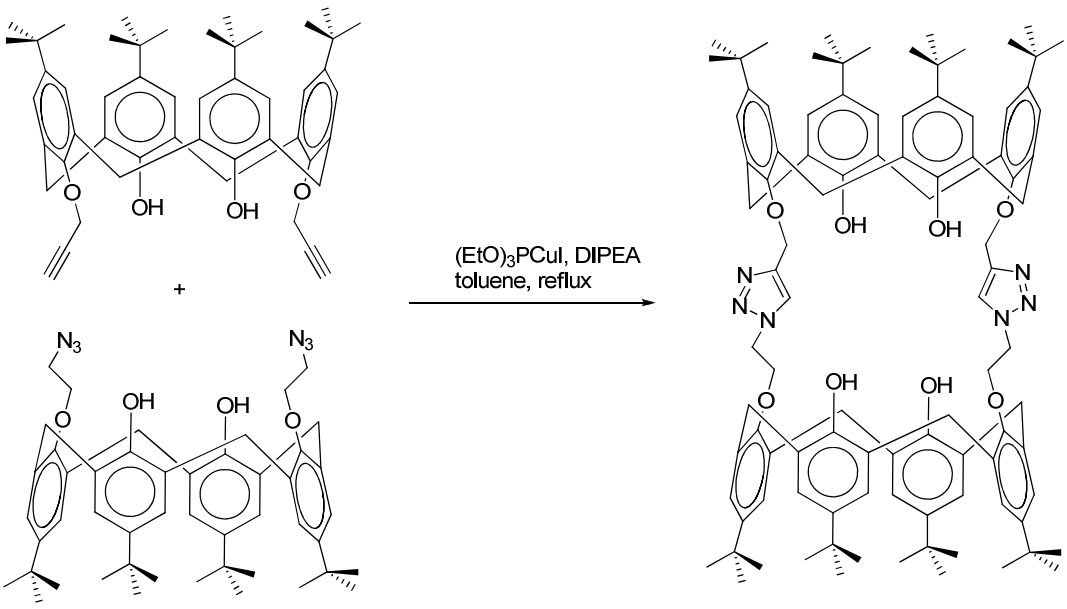
"Click chemistry" has been used in applications of nanoparticle targeting *in vivo*[66]. Cyclic peptide Lyp-1 (CGNKRTGC) is a known tumor targeting peptide. It selectively binds to p32, a mitochondrial proteins that is both overexpressed and aberrantly localized at the cell surface of tumor cells, macrophages and lymphatic endothelial cells in certain experimental tumors and in human cancers. With the application of Hüisgen cycloaddition, alkyne-bearing Lyp-1 was treated with azido-PEG bearing iron oxide nanoparticles in the presence of Cu(I) (resulting in Lyp-1 conjugated nanoparticles (**Scheme 4.8**)). After injection into mice bearing human cancer cells, the nanoparticles were found to accumulate at the tumor site.



Scheme 4.8: Hüisgen cycloaddition applied for Cellular targeting nanoparticles.

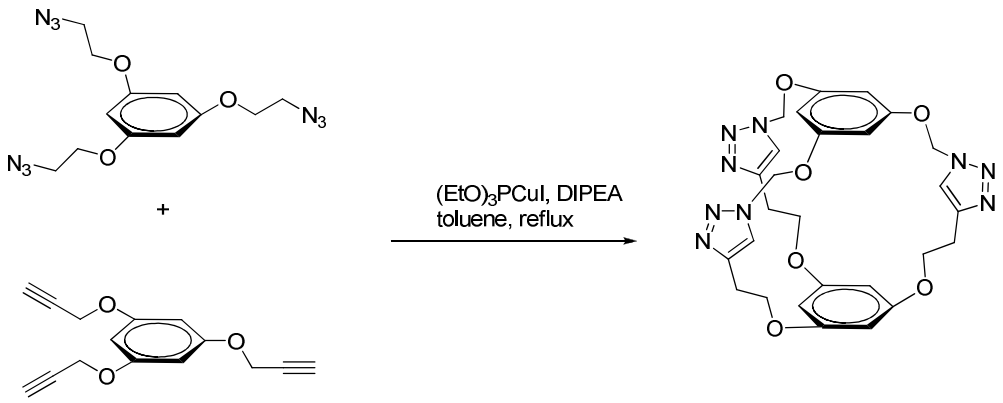
4.1.4.3 Synthesis of nanotubes and nanocages.

Hüisgen cycloaddition has been shown to be useful in creating molecular architectures. For instance, starting from calixarene-derivatives, cycloaddition results in nanotubes that can entrap gaseous molecules and metal ions. Double Hüisgen cycloaddition of dipropargylate calix[4]arene and diazide calix[4]arene yielded bis-calix[4]arene nanotube in the presence of Cu(I) catalysts (**Scheme 4.9**)[(67)].



Scheme 4.9: Synthesis of bis-calix[4]arene nanotube by mean of Hüisgen cycloaddition.

By using similar reaction conditions, structures called molecular nanocages were achieved from polyalkyne and polyazide compounds (**Scheme 4.10**)[68]. Once the first cycloaddition took place, the following reactions occured in an intramolecular manner. Polymerisation due to cross cycloaddition could be avoided by performing the reaction at high dilution.

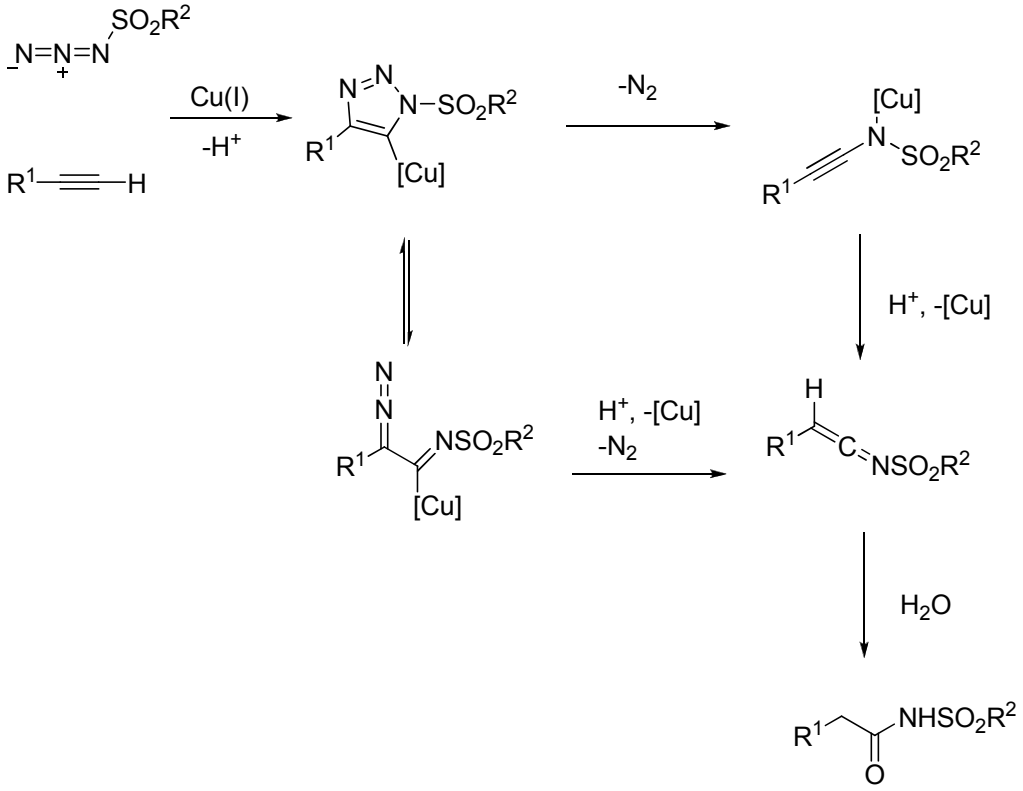


Scheme 4.10: Synthesis of nanocages by means of Hüisgen cycloadditions.

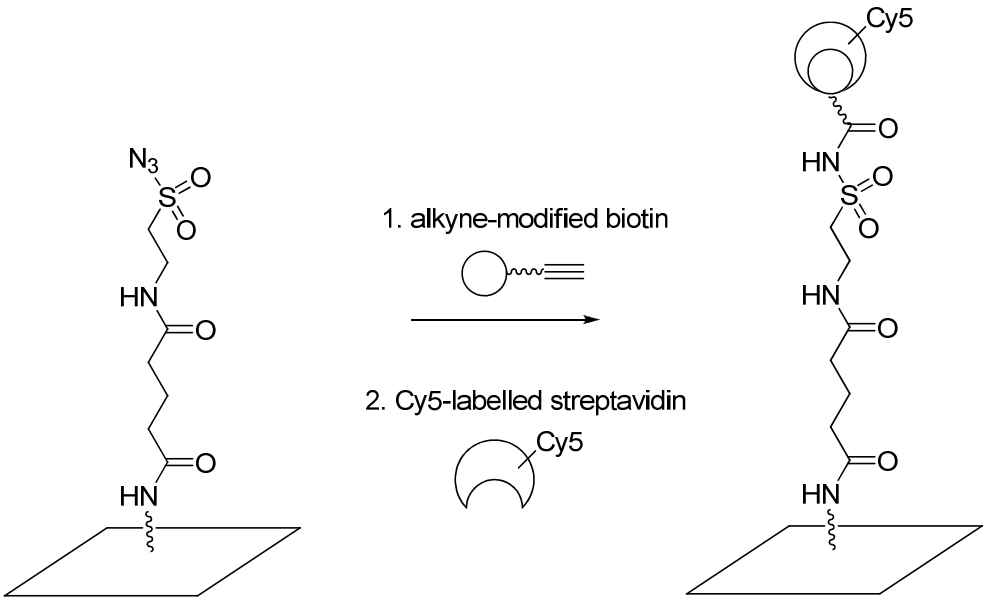
4.1.4.4 Surface immobilisation by click sulfonamide reactions.

The "Click sulphonamide" reaction is a modified Hüisgen 1,3-dipolar cycloaddition in which sulfonylazides are used as dipolarphiles. Following liberation

of N_2, the "Click sulphonamide" results in *N*-acylsulfonamides instead of 1,2,3-triazoles. **Scheme 4.11** represents a proposed mechanism of this formation[69]. This reaction has been applied to allow immobilization of biomolecules onto glass slides. Following functionalisation of a glass slide with sulfonylazide, alkyne-modified biomolecules, including biotin, sugar and peptides were introduced and these biomolecules were covalently bound onto the glass slide (**Scheme 4.12**)[70].



Scheme 4.11: Proposed mechanism of *N*-acylsulfonamide formation *via* Hüisgen 1,3-dipolar cycloaddition between alkynes and sulfonylazides.

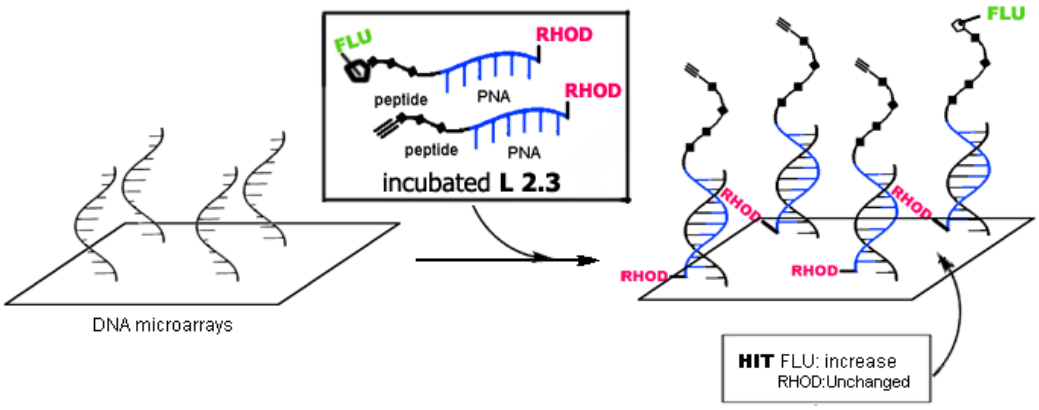


Scheme 4.12: Immobilization of biomolecules onto sulfonylazidefunctionalized glass slides.

4.2 Result and Discussion.

4.2.1 The Concept:

The experimental design for the search of peptide sequences as potential substrates for an internally promoted Hüisgen cycloaddition are summarised in **Scheme 4.13**. A PNA-encoded peptide library was generated *via* split-and-mix synthesis on PEGA resin. Each position of this tetrapeptide library was varied up to 6 amino acids including D-Lys, D-Glu, D-Tyr, D-Pro, D-Ala and Gly and each amino acid was encoded with a PNA quartet. The library was functionalised with an alkyne on the amino terminus of the peptide chain and rhodamine B was coupled onto the PNA chain. After being cleaved from resin, the library was incubated with 5-azidofluorescein. The library was then hybridised onto a DNA microarray and the "hit" members were then identified by fluorescence scanning. A control was performed under the same conditions except that copper iodide was added into the reaction as a catalyst.

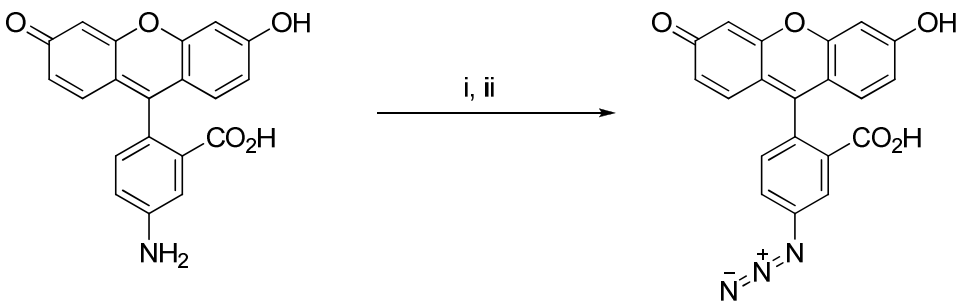


Scheme 4.13: Experimental procedure for screening for "peptide based" Hüisgen cycloaddition catalyst.

4.2.2 Hüisgen cycloaddition of an alkyne modified-peptide on a solid support with 5-azidofluorescein.

In order to evaluate the Hüisgen cycloaddition between peptide sequences and 5-azidofluorescein and generate a control a *N*-5-hexynoyl-modified tetrapeptide of the sequence Ala-Gly-Ala-Gly was synthesised and reacted with the azido fluorophore.

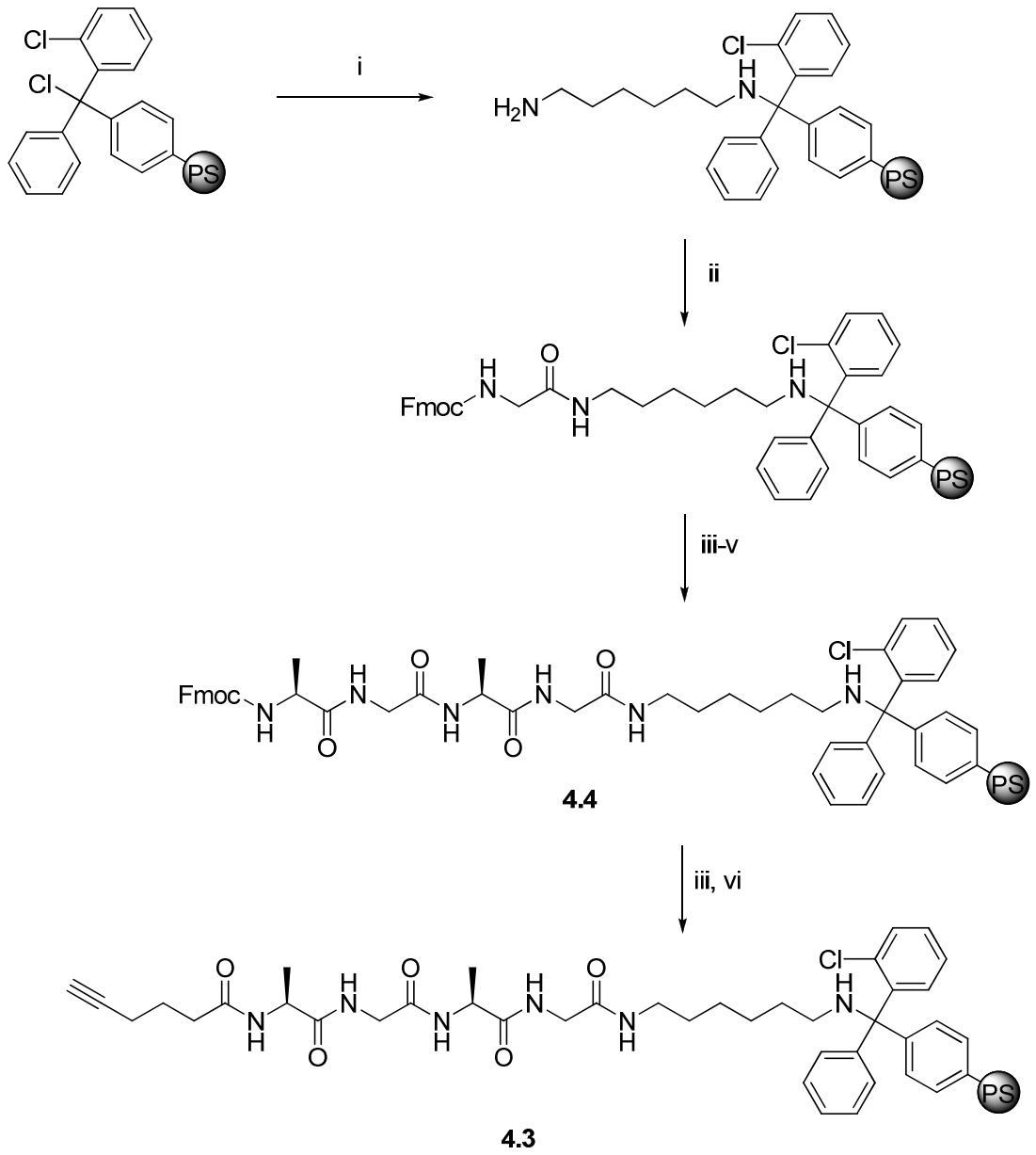
5-Azidofluorescein (**4.2**) was prepared from 5-aminofluorescein (**4.1**) by a Sandmeyer reaction (**Scheme 4.14**)[71].



4.1 **4.2**

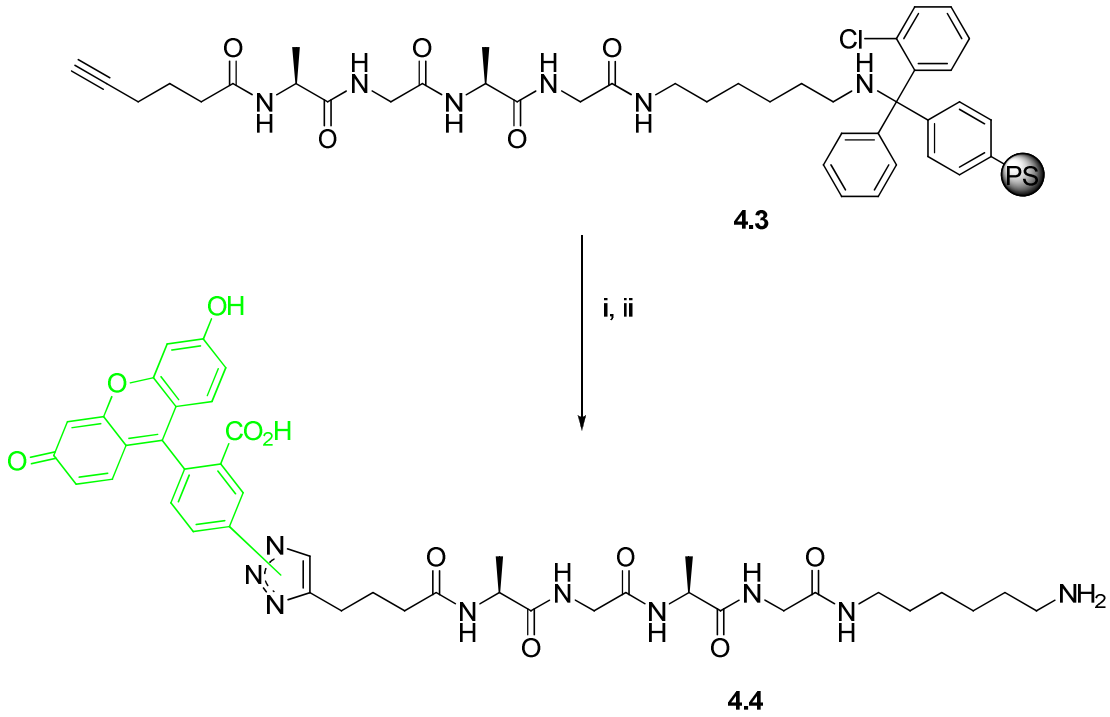
Scheme 4.14: Preparation 5-azidofluorescein; *Reagents and conditions*: (i) $NaNO_2$,HCl, 0°C; (ii) NaN_3.

N-5-hexynoyl-modified Ala-Gly-Ala-Gly (**4.3)** was synthesised as shown in **Scheme 4.15** using chlorotrityl-resin. After the coupling of 1,6-diaminohexane onto chlorotrityl resin, Fmoc-protected amino acids were coupled under microwave irradiation, giving tetrapeptide **4.4**. After the Fmoc deprotection, 5-hexynoic acid was coupled using microwave irradiation, giving resin **4.3**.



Scheme 4.15: Synthesis of *N*-5-hexynoyl-modified Ala-Gly-Ala-Gly on chlorotrityl resin (**4.3**); *Reagents and conditions:* (i) 20 equiv. 1,6-diaminohexane, DCM, 2 h; (ii) Fmoc-Gly-OH; (iii) 20% piperidine/DMF, 2 x 10 min; (iv) Fmoc-Ala-OH or Fmoc-Gly-OH; (v) repeat (iii),(iv) as necessary; (vi) 5-hexynoic acid; All coupling reactions employed the following conditions: 5.0 equiv. Amino acids or 5-hexynoic acid, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, μw irradiation.

Resin **4.3** was reacted with 5-azidofluorescein (**4.2**) in the presence of copper iodide and a catalytic amount of *N,N*-Di*iso*propylethylamine (DIPEA) (**Scheme 4.16**).



Scheme 4.16: Hüisgen cycloaddition of *N*-5-hexynoyl-modified Ala-Gly-Ala-Gly on solid support; *Reagents and conditions*: (i) 1.8 equiv. 5-azido fluorescein (**4.2**), 1.4 equiv. CuI, cat. DIPEA, DMF, 8 h; (ii) 1:1 TFA:DCM, 1 h.

After 8 h. of reaction, excess reagents were washed away; the peptide sequence was cleaved from the solid support and analysed by analytical HPLC and mass spectrometry (**Figure 4.1**).

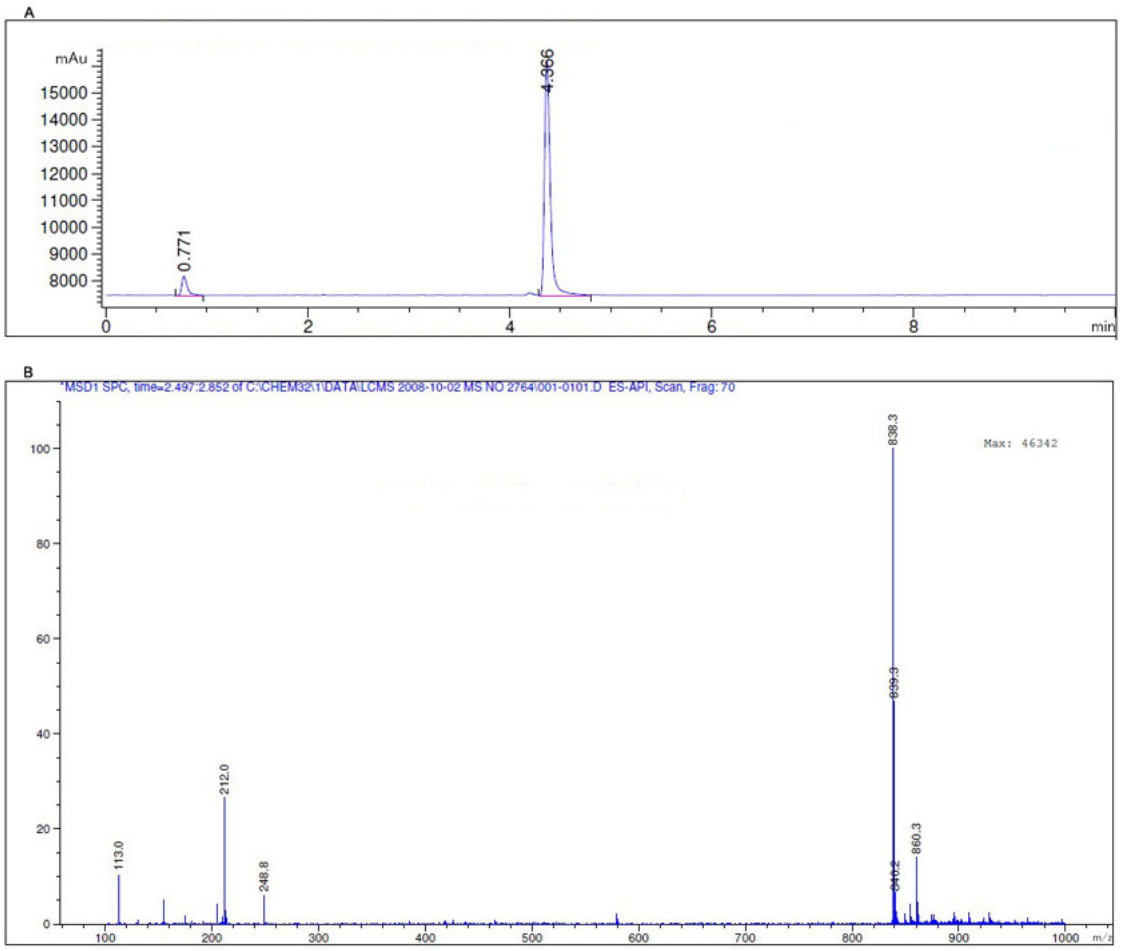


Figure 4.1: HPLC trace (**A**) and mass spectrum (negative scan) (**B**) of product from reaction between **4.2** and **4.3**.

Figure 4.1 confirmed the triazole **4.4** as the major product in the reaction. HPLC analysis showed a single peak whose mass corresponded to the expected product (found *m/z* = 838.3 $[M-H]^-$, calculate *m/z* = 838.3 $[M-H]^-$).

As a negative control, 5(6)-carboxyfluorescein was reacted with **4.3**. The reaction was treated identically to the reaction between **4.2** and **4.3**, and HPLC analysis and mass spectrum of reaction mixture are shown in **Figure 4.2**, representing mass corresponding to cleaved **4.3**; Therefore no reaction tookplace in the control.

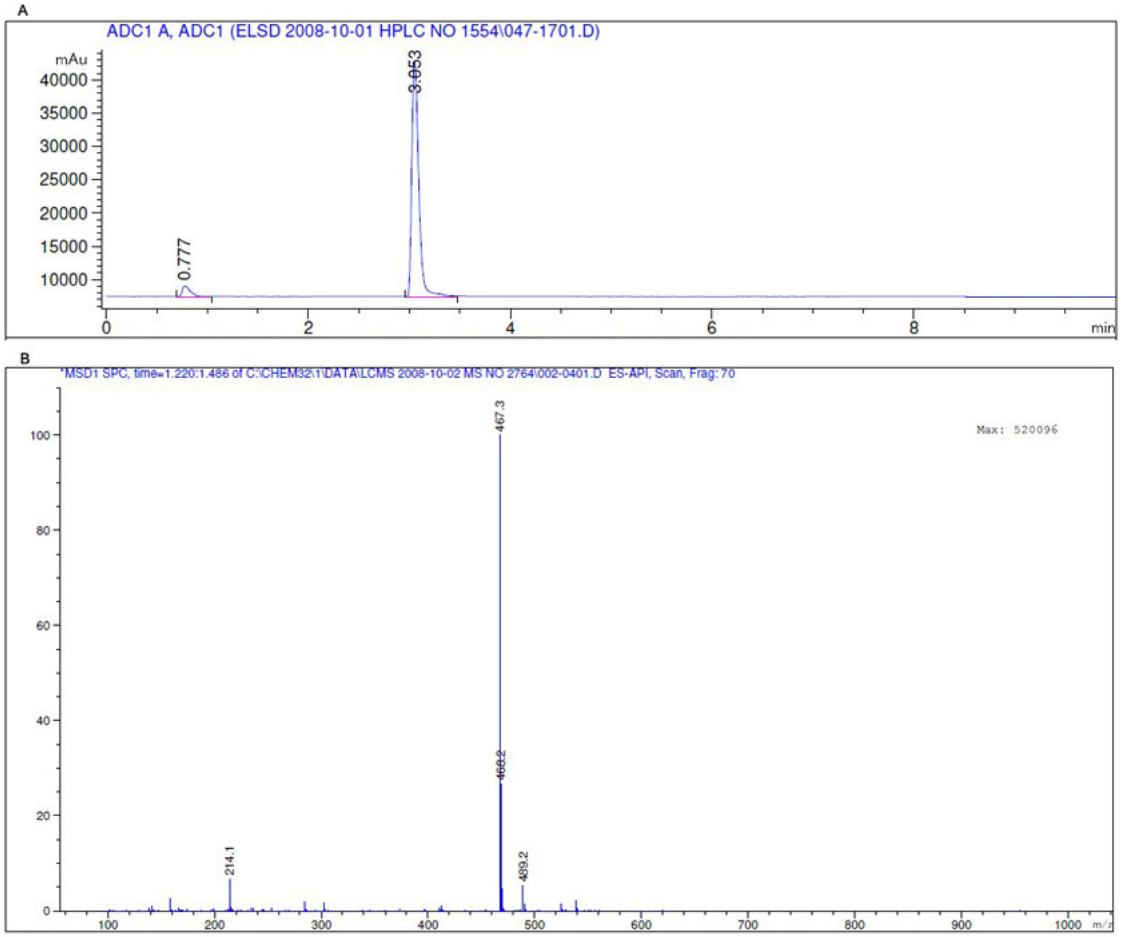


Figure 4.2: HPLC trace (**A**) and mass spectrum (positive scan) (**B**) of the product from reaction between **4.2** and 5(6)-carboxyfluorescein.

In conclusion, the alkyne-functionalized peptide and azidofluorescein could be used as the components in Hüisgen cycloaddition.

4.2.3 Screening of a peptide library as potential substrates for internally promoted Hüisgen cycloaddition.

100 μM of library **L 2.3**, a peptide library functionalized with an alkyne and labelled with rhodamine B, in aqueous solution was mixed with 200 μM of 5-azidofluorescein and the reaction mixture was left overnight. Unreacted 5-

azidofluorescein was removed by a centrifugal filter device (Millipore) and the library was subsequently hybridised onto a DNA microarray.

Two controls were also performed. As a positive control (control 1), 100 μM (final concentration) of copper iodide was also added to the mixture and for a negative control (control 2), only the library solution was hybridised onto DNA microarrays.

Microarrays were scanned using a fluorescence scanner using 2 filter sets in order to detect two wavelengths corresponding to fluorescein and rhodamine B. In addition, a modified set-up was also used. In this set-up, arrays were excited at 490 nm (fluorescein maximum excitation wavelength) and emission detected at 580 nm (rhodamine B maximum emission wavelength) because "hit" members would result in the introduction of fluorescein onto the peptide, enabling FRET.

"Hit" members from this screening procedure resulted in introduction of fluorescein onto molecules; they therefore could be identified by high fluorescein intensity. However, the high fluorescein intensity on the spots could be the results of other effects such as hybridisation efficiency. As RHOD was the internal control that existed within every library member (**Scheme 4.13**) and its intensity was not changed under the reaction conditions, the FLU/RHOD ratios were therefore taken into consideration instead of absolute fluorescein intensities. As a result, higher FLU/RHOD ratios would identify the "hit".

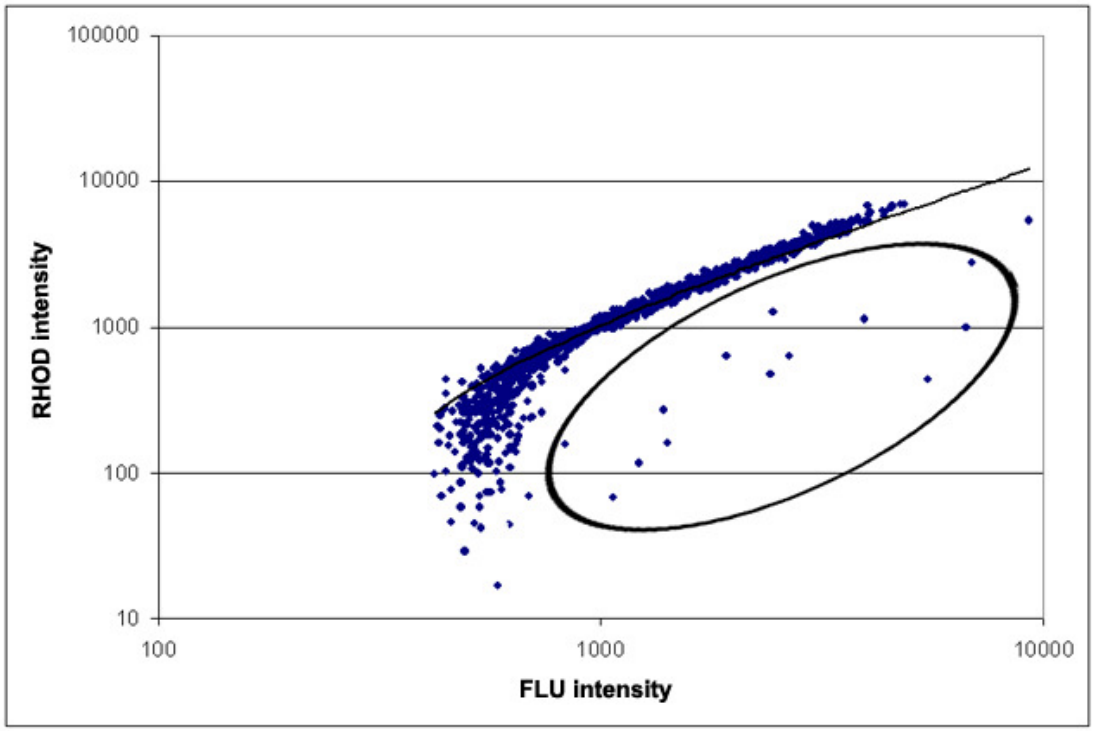


Figure 4.3: Ratio of fluorescent intensities obtained from library **L 2.3**, treated with 5-azidofluorescein.

Figure 4.3 shows the fluorescent intensities of library **L 2.3**, reacted with 5-azidofluorescein. Both FLU and RHOD intensities were seen to vary in a wide range. However, most members showed relatively constant FLU/RHOD ratio (spots distributed on diagonal line). According to the experimental design, "hit" members are expected to show higher FLU/RHOD ratios due to the introduction of fluorescein to their peptide ends. The spots representing these peptides can be identified from **Figure 4.3**. These spots show the higher FLU/RHOD ratios than the average, and they are shown in the oval.

Data from **Figure 4.3** was reploted as FLU/RHOD ratios against peptide sequences (**Figure 4.4**). It was clearly shown that, while most of library members have constant FLU/RHOD ratio at about 1.0, there are some members that showed significantly higher ratios.

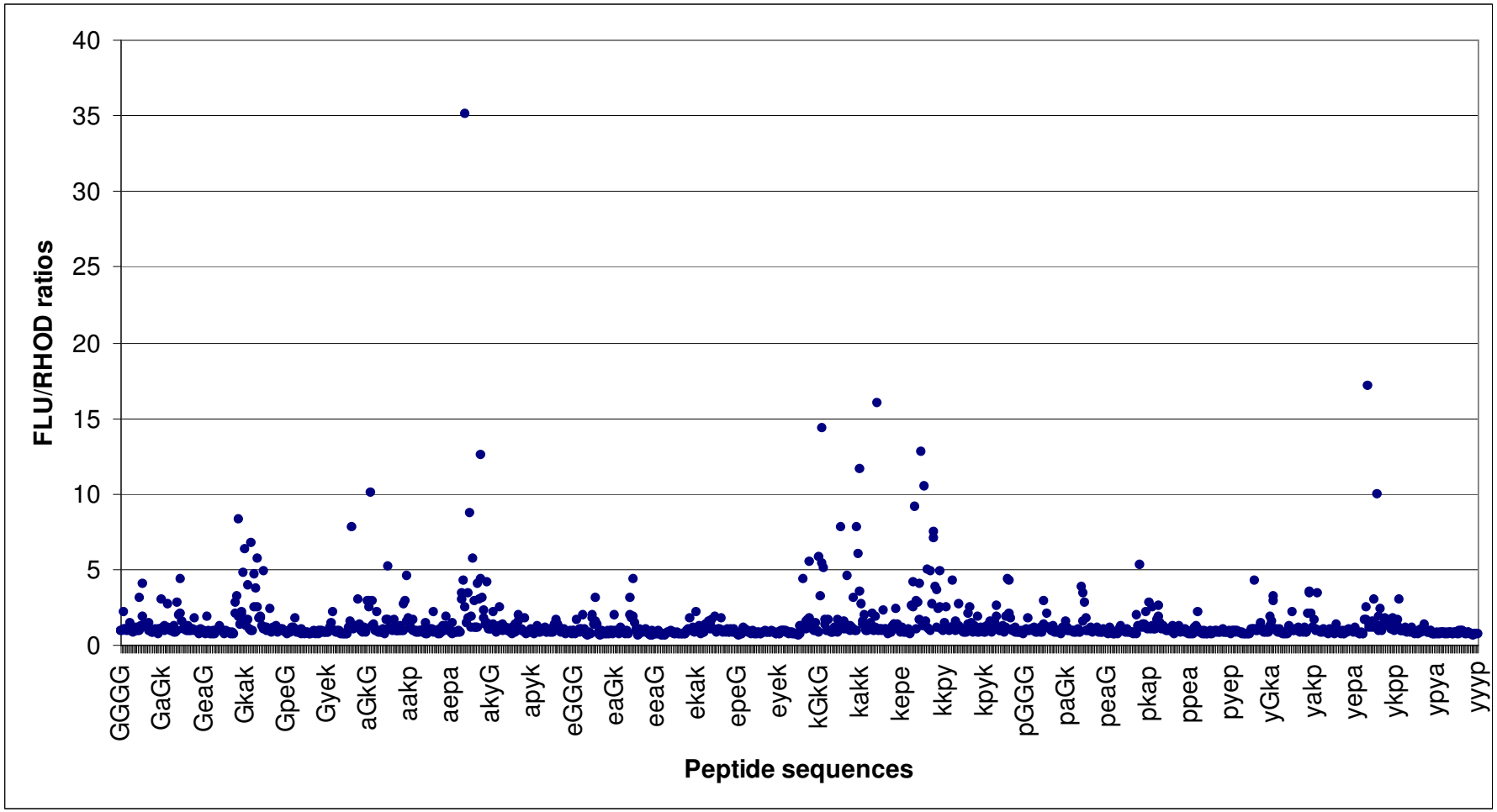


Figure 4.4: FAM/RHOD ratio of library **L 2.3** treated with 5-azidofluorescein.

Data with FLU/RHOD ratios higher than 2.0 were selected and the amino acid sequences were analysed by Spotfire®. From 135 sequences showing the highest FLU/RHOD ratios, D-Lys was a preferable amino acid at position AA^1, AA^2 and AA^3, D-Lys and D-Ala were preferable at position AA^4. In addition, no D-Tyr was found at position AA^1 of this group (**Figure 4.5**).

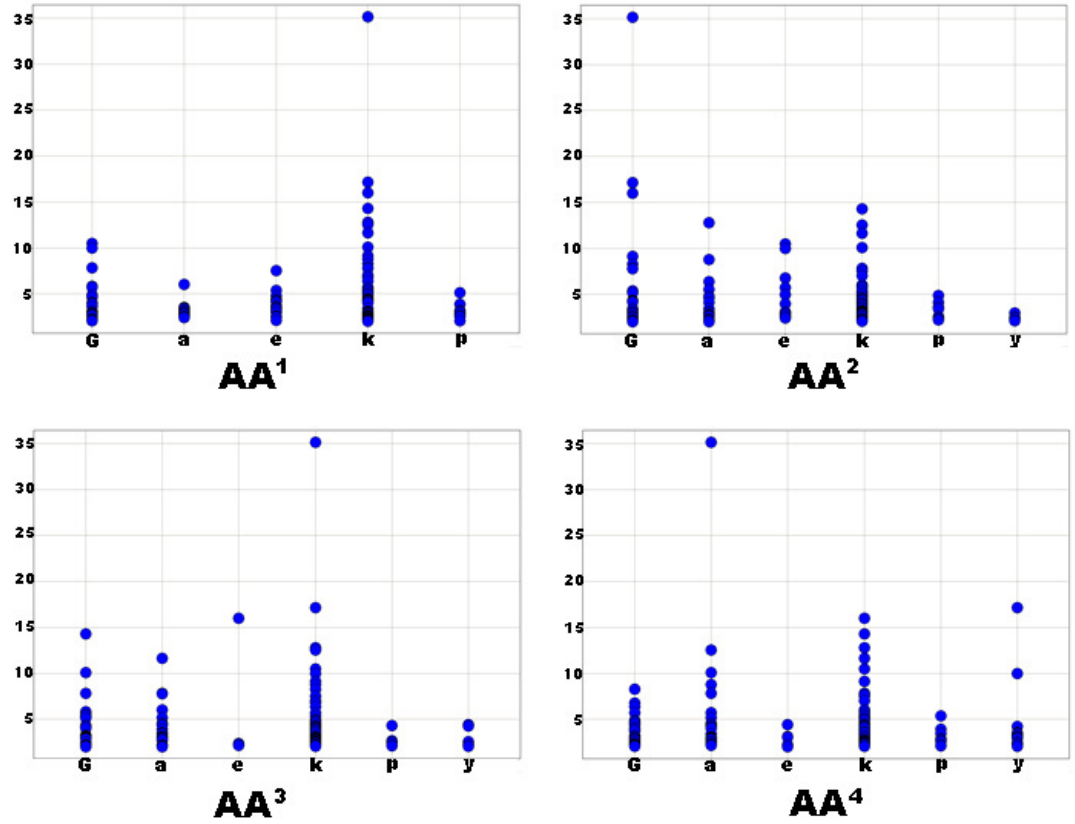


Figure 4.5: Representation of the top 135 peptides potentially capable of promoting Hüisgen cycloaddition. Each panel represents the presence of a certain amino acid at each position.

However, control 1 showed very similar results despite the fact that one equivalent of copper iodide was added into the reaction. FLU/RHOD ratios against peptide sequences from this control (**Figure 4.6**) showed a very similar pattern as that in **Figure 4.4**. By sequence analysis (Spotfire®), similar results to those in the experiment were obtained (data not shown).

Due to the difficulty of CuI solubilization in water, as well as disproportionation of copper(I), giving copper(0) and copper(II), it is possible that, in control 1 (the reaction mixture containing CuI), the copper(I) available in the reaction mixture was not sufficient. As a consequence, cycloaddition taking place in control 1 might be the result of the peptides' ability to "help" the reaction.

Control 2 showed the same FLU/RHOD ratios for every library member (**Figure 4.7**), confirming that the increase in the fluorescein intensity as shown in **Figure 4.6** was not random.

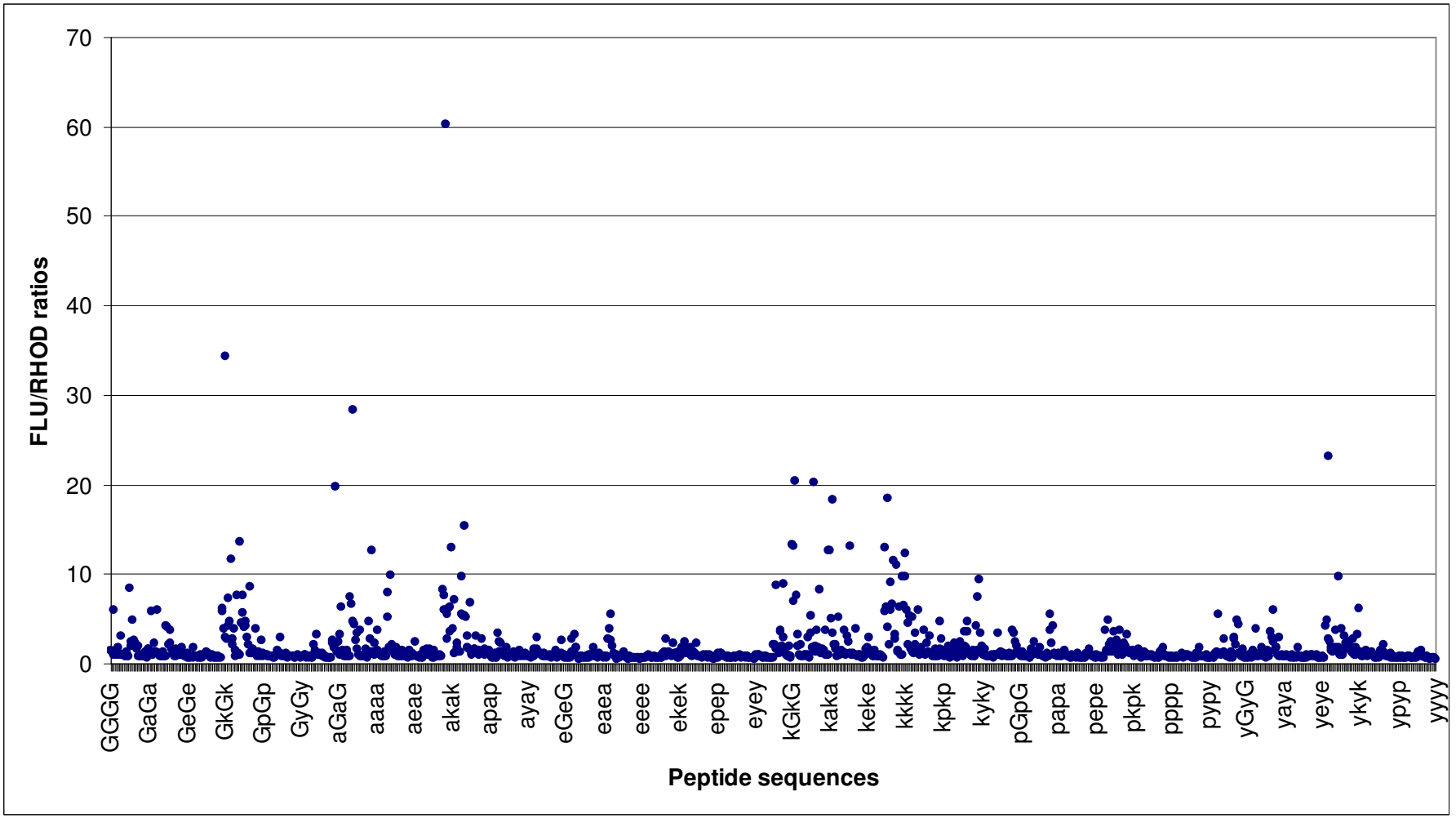


Figure 4.6: FAM/RHOD ratio of library **L 2.3** treated with with 5-azidofluorescein and copper iodide (control 1).

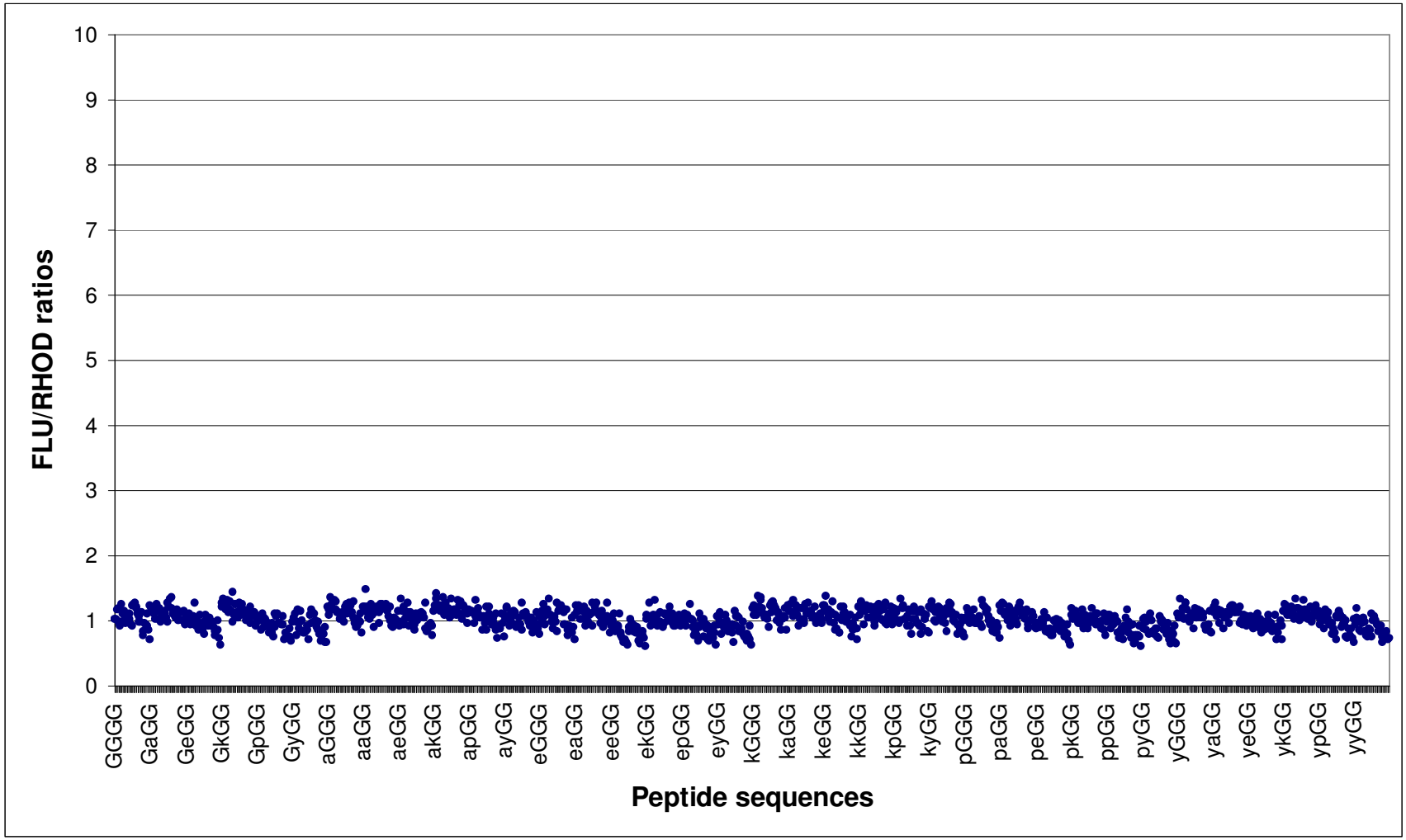


Figure 4.7: FAM/RHOD ratio of library **L 2.3** (control 2).

FRET intensities obtained from library **L 2.3** reacted with 5-azidofluorescein (**Figure 4.8**) were also analysed. “Hit” members resulted in an introduction of fluorescein onto the peptides and therefore acting as donor chromophore so when the microarrays were excited at 490 nm and emission would be detected at 580 nm with high intensities expected from the peptide with fluorescein conjugated. From the data presented in **Figure 4.8**, spots with highest FRET intensities were selected for amino acid sequence analysis (**Figure 4.9**). From the150 members that showed the highest FRET intensities, D-Glu, D-Pro and D-Tyr were found to be preferable at every position.

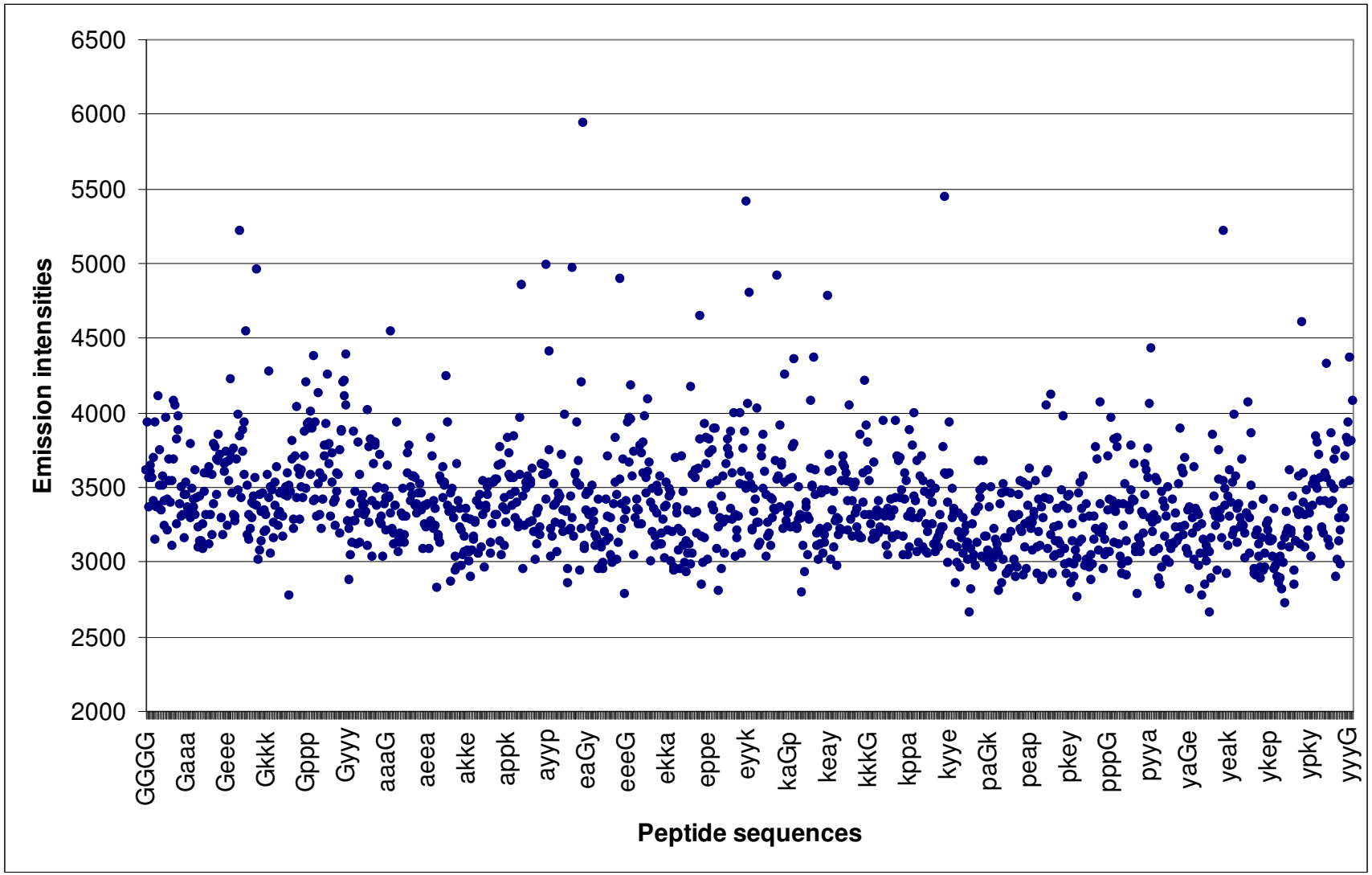


Figure 4.8: FRET intensity of library **L 2.3** treated with 5-azidofluorescein.

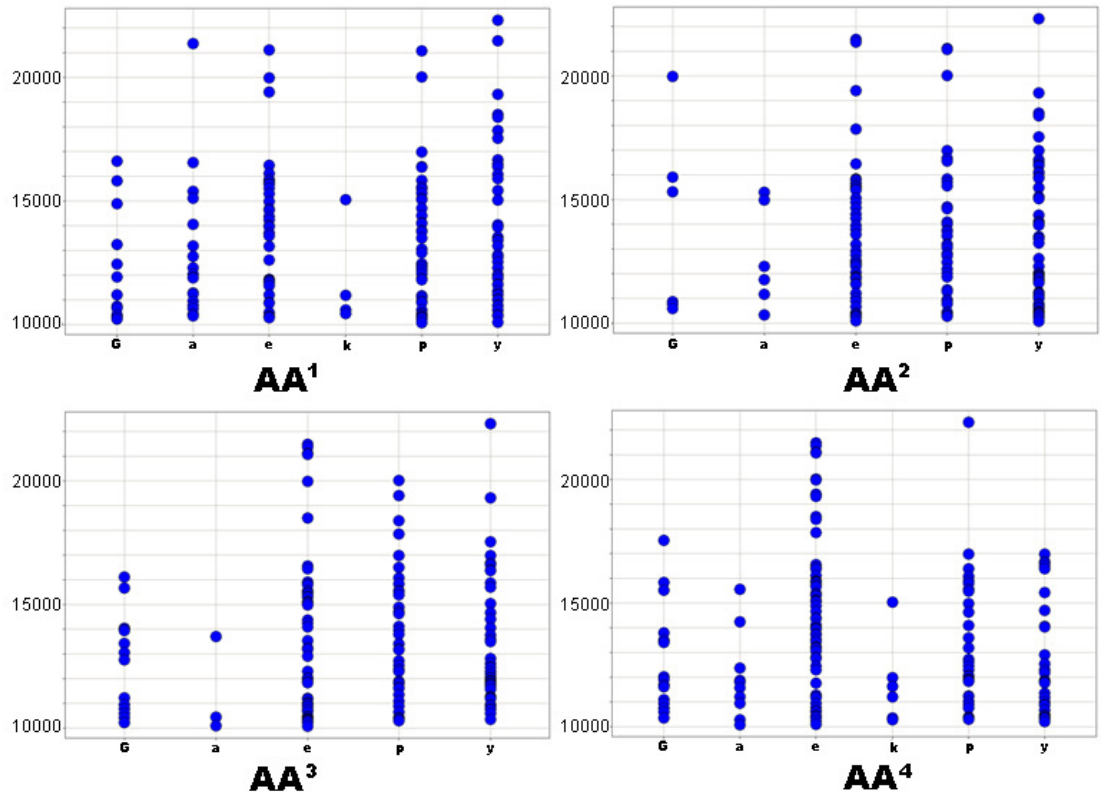


Figure 4.9: Representation of the peptides sequences showing high FRET intensities after Hüisgen cycloaddition. Each panel represents the presence of a certain amino acid at each position.

It appears that different data analysis methods gave contradictory results. While comparison of FLU/RHOD ratio indicated that D-Lys was a preferable amino acid, analysis of FRET intensities indicated that D-Glu, D-Pro and D-Tyr were preferable and D-Lys was the least preferable amino acid. For the FRET phenomenon, in addition to the existence of suitable chromophores on the analysed peptide sequences, the distances between both chromophores is very important. In library **L 2.3**, rhodamine B and the alkyne group are located far away from each other. If the cycloaddition took place, introducing fluorescein onto the peptide amino terminus, the FRET effect might not take place depending on the peptide conformation. In case of a library like **L 2.3**, in which chromophores are certainly not

in the effective range for FRET, comparison of emission intensity from each chromophore is the more reliable analysis method.

4.2.4 Synthesis and evaluation of a "hit" peptide sequence as a substrate for internally promoted Hüisgen cycloaddition

From data analysis described in 2.4.3, the sequence D-Ala-D-Lys-Gly-D-Lys was identified as a "hit". The *N*-5-hexynoyl-modified sequence was synthesised on aminomethyl-polystyrene resin linked by a Rink amide linker (see **Figure 2.9**). *N*-5-hexynoyl-D-Ala-D-Lys-Gly-D-Lys-NH-Rink-PS (**4.5**, **Figure 4.10**) and was subjected to the solid-phase reaction with azidofluorescein (**4.2**) in DMF – the reaction conditions identical to those used in **4.2.2** except that the reaction time was longer (16 h). Such reaction is presented in Entry **1** of **Table 4.1**, while the positive control, where CuI was added to the reaction mixture (Entry **2** of **Table 4.1**), was also carried out.

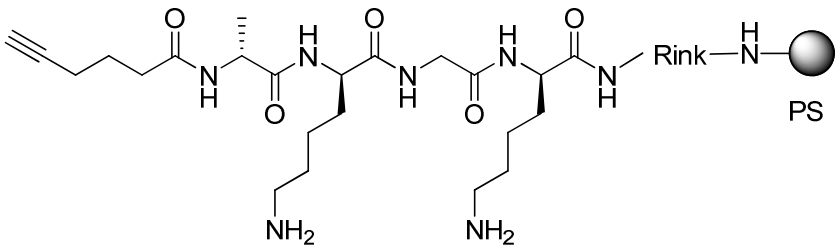


Figure 4.10: *N*-5-hexynoyl-D-Ala-D-Lys-Gly-D-Lys-NH-Rink-PS (**4.5**).

Entry	Reaction mixtures
1	**4.5** + 5-azidofluorescein (**4.2**, 1.8 equiv.)
2	**4.5** + 5-azidofluorescein (**4.2**, 1.8 equiv.) + CuI (1.4 equiv.) + DIPEA (2 drops)

Table 4.1: Reaction mixtures in the evaluation of "hit" peptide sequences as a substrate for internally promoted Hüisgen cycloaddition, using **4.5** as substrate. *Reaction conditions:* 16 h. in DMF.

After leaving the reactions overnight, the solid support was washed and the peptide sequences were cleaved. The resultant residue was analysed by HPLC and mass spectroscopy. HPLC analysis (**Figure 4.12**) showed that the reaction mixture from entry **1** did not show the existence of any cyclo-adduct (**4.6**, **Figure 4.11**). The major compound found in the mixture was starting material. The product from the cycloaddition could be detected only in reaction from entry **2**, where the Cu(I) catalyst was used. In this reaction, the new peak at 2.9 min was clearly detected, and MS analysis showed a mass corresponding to the expected product (**4.6**) (found *m/z* = 869.3 $[M+H]^+$, calculate *m/z* = 869.4 $[M+H]^+$). From these results, it could be concluded that the sequence identified as the "hit", D-Ala-D-Lys-Gly-D-Lys, did not promote an internal Hüisgen cyclo-addition, despite appearing to show this on the array when PNA-encoded.

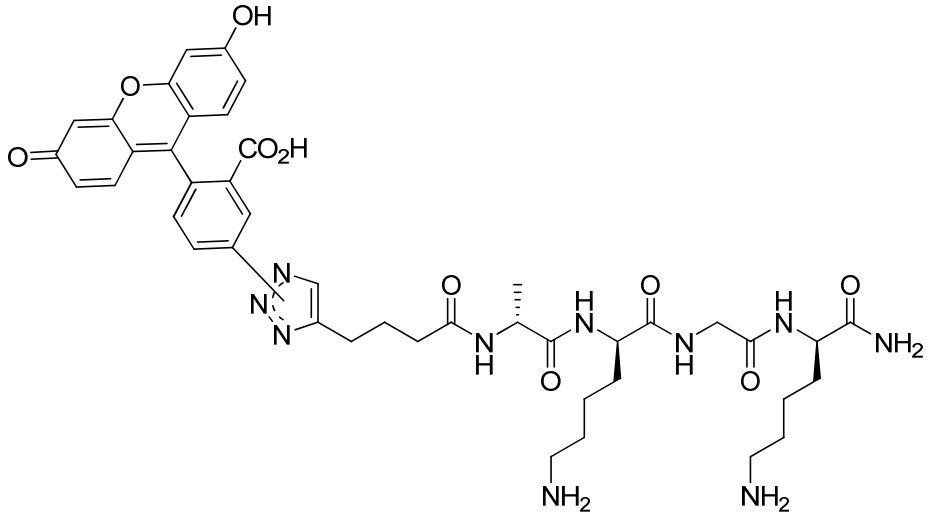


Figure 4.11: Compound **4.6**, an expected product (after cleavage) from the cycloaddition between solid support bearing peptide (**4.5**) and 5-azidofluorescein (**4.2**).

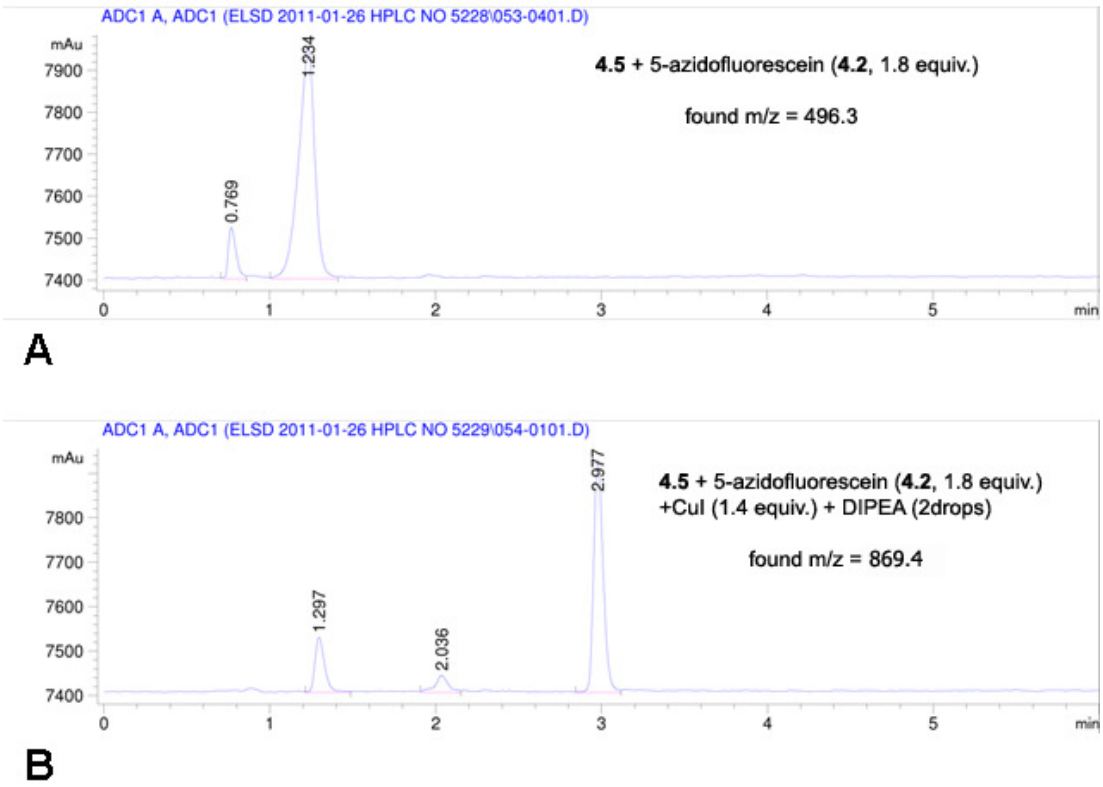


Figure 4.12: HPLC analysis of reaction between **4.5** and 5-azidofluorescein without Cu(I) catalyst (**A**) and with Cu(I) catalyst (**B**).

4.3 Conclusions.

In conclusions, a library of alkyne functionalized PNA-encoded D-amino acid-containing peptide was screened for potential substrates of self-promoting Hüisgen cycloaddition. Using this method the tetrapeptide D-Ala-D-Lys-Gly-D-Lys was identified as a peptide potentially capable of promoting a Cu(I) free Hüisgen cycloaddition, however, the evaluation of the self-promotion Hüisgen cycloaddition of such the sequence proved negative when scaled-up. This was unexpected because the signal of the "hit" peptide was unambiguous (FLU/RHOD ratio was as high as 35).

The results can probably be explained that the reaction conditions used in the screening of the library and in the evaluation of the "hit" sequence were different, which could affect the properties of the peptide. While the screening of the library took place in solution, the reaction of the "hit" sequence took place in solid phase. This difference might affect the accessibility of the function group on the peptide sequences. In addition, the library was screened in aqueous solution while the reaction of the "hit" was in DMF due to the resin used and the need to solubilise the 5-azidofluorescein.

Having been failed in the Cu(I) free Hüisgen cycloaddition using the "hit" peptide sequence as substrate, the experimental procedures and data analysis were reinvestigated. One issue was that the library screening was not repeated to confirm the results (n=1). This is usually avoided because of the issue of false positives, but arose due to time constraints. Another possible mistake was that the screening of the library was carried out in aqueous solution while it should have been in a buffer solution. As a result, a number of library members were found to have decreased rhodamine B (the control) intensities after the incubation, hence mistakenly identified as "hit" due to the high FLU/RHOD ratios.

This experiment can probably be improved by doing reaction on microarray, in which library **L 2.3** is hybridised prior to the treatment of azidofluorescein (**4.2**). By doing so, the intermolecular catalyzed reaction – where one peptide chain catalyzes cycloaddition on another peptide – can be avoided. Since the screening of this library used fluorescein derivative as a probe, it is possible that the lactone isomer of the dye was attacked by peptide's nucleophilic side chain e.g. Lys, leading to the false positive. This problem can be avoided by using alternative dye lacking of

carboxylic acid such as azido derivative of Tokyo Green. In conclusions, the screening of library **L 2.3** and the "hits" identified may be artifactual due to the design of the experiment.

Chapter 5: High-throughput screening of a PNA-encoded peptide library for the discovery of cell-penetrating peptides.

5.1 Introduction.

5.1.1 Structure of cell-penetrating peptides.

Cell-penetrating peptides (CPPs) are short peptides sequences (less than 30 amino acids) capable of translocating across the cell membrane[72]. There are a wide variety of existing CPP structures, however, they commonly contain positively charged amino acids, such as, lysine or arginine residues (**Table 5.1**).

Name	**Sequence**
TAT peptide (48-60)[73]	GRKKRRQRRRPPQ
Penetratin[74,75]	RQIKIWFQNRRMKWKK
Peptide-based gene delivery systems (MPG) [76,77]	GALFLGFLGAAGSTMGAWSQPKKKRKV
Transportan[74,75]	GWTLNSAGYLLGKINLKALAALAKKIL
Polyarginine CPP[78,14]	R_8

Table 5.1: Examples of CPP.

5.1.1.1 TAT peptide (48-60).

TAT is an 86-amino acid protein involved in the replication of human immunodeficiency virus type 1 (HIV-1). This protein is able to translocate through the plasma membrane and reach the nucleus. The domain responsible for its cell

penetrating property was determined by evaluating the cellular delivery of derivatives of TAT-(37-60), with a sequence consisting of an α-helical domain (residues 37-47) and a basic domain (residues 48-60). The basic domain was found to play a major role in cellular delivery as an incomplete this domain resulted in the reduction of cell internalization efficiency[79].

5.1.1.2 Penetratin.

Penetratin is part (residues 43–58, RQIKIWFQNRRMKWKK) of a protein domain that binds DNA or RNA and which first discovered in a genus of small flies (*Drosophila*)[74,75]. In this peptide, 7 of the 16 residues are positively charged (Lys and Arg). It was found that a 16 amino acid sequence was the minimum length required for internalization to nerve cells. Shorter sequences were not found to have this property[80].

5.1.1.3 Peptide-based gene delivery systems.

The peptide vector of the sequence GALFLGFLGAAGSTMGAWSQP KKKRKV was used in the delivery of oligonucletides into cells. It consists of hydrophobic and hydrophilic residues and rapidly delivers single and double stranded oligonucleotides into mammalian fibroblast cells in less than 1 h[76]. Seventeen amino acids at the amino terminus of this vector (GALFLGFL GAAGSTMGA) are responsible for efficient movement across the cell membrane. The KKKRKV sequence is required for interaction with the nucleic acids and a WSQP sequence functions as the linker between the two sequences[77]. The peptide

vector binds to oligonucleotides and the mixture forms complexes which help to increase the stability of the oligonucleotides by limiting access to nucleases[76].

5.1.2.4 Transportan.

Transportan is a 27 amino acid peptide whose sequence was modified from the known CPP, galparan (GWTLNSAGYLLGPINLKALAALAKKIL)[81]. Galparan was originally designed from the combination of 12 amino acids from the amino terminus of the neuropeptide galanin (GWTLNSAGYLLGPHIDNHRSFHD KYGLA) and mastoparan a wasp venom peptide toxin of the sequence INLKALAALAKKIL. Mastoparan is responsible for cell penetration while the amino terminal part is a galanin receptor ligand with agonistic properties. Transportan has been modified by replacement of proline by lysine and it was found to be delivered into Bowes' melanoma cells within 1 minute[82].

5.1.1.5 Polyarginines.

The TAT peptide (48-60) has 6 amino acids that are arginine (out of 13). This amino acid is therefore believed to play a critical role in cell internalizing processes. Polyarginine peptides were therefore evaluated as cellular delivery agents and it was found that a protein with a size of 29 kDa was successfully delivered into mouse macrophages cells by the R_9 peptide. After optimization of sequence length, R_6 and R_8 were found to be the most efficient sequences for cellular delivery[78]. R_3 and R_7-PNA conjugates were also synthesised and these polyarginines successfully delivered PNA oligomers into cells[14].

5.1.2 Cell-penetrating peptides - cellular uptake mechanisms.

The cell membrane (also called the plasma membrane, or phospholipid bilayer) is a physical boundary found in all types of cell responsible for isolating the cell content from the extracellular environment and controlling efflux/influx of solutes from cells. Its structure contains a wide variety of biological molecules including lipids, proteins and carbohydrates; however, the major component is phospholipids-based. A phospholipids is an amphipatic molecule consisting of a hydrophilic polar head in the form of a negatively charged phosphate group, and two hydrophobic fatty acid tails. To form the membrane, phospholipids are held together by non covalent interactions and arranged in a continuous double layer called the lipid bilayer (**Figure 5.1**).

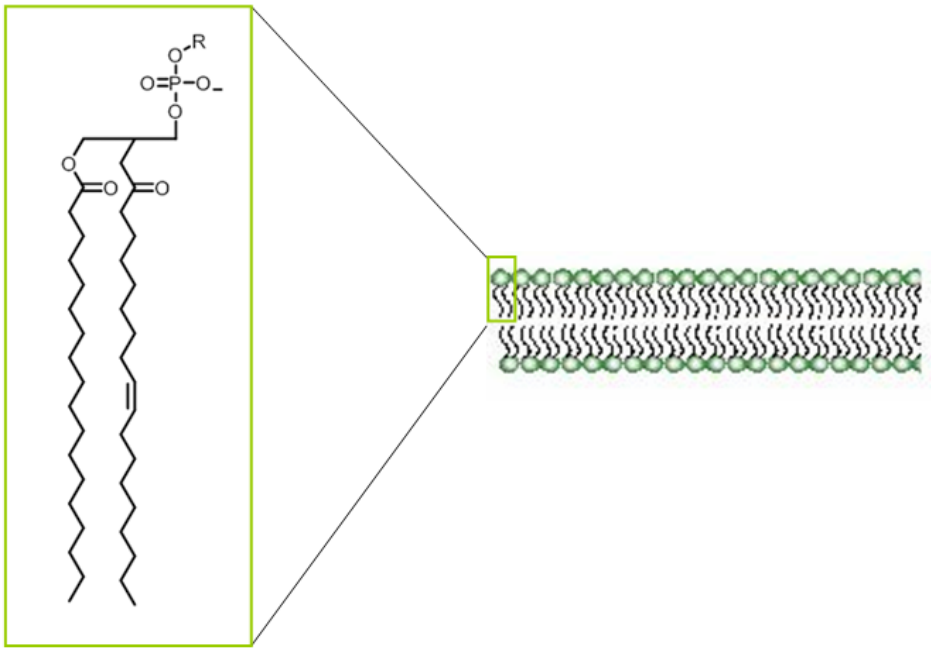


Figure 5.1: Phospholipid structure and cell membrane.

CPPs can access inside the cells by two distinct methods, energy-dependent endocytosis and energy-independent direct translocation across cell membrane[83]. In

endocytosis, CPPs internalize into cells by encapsulation into lipid vesicles (liposomes), which are internalized after resealing of the cell membrane. Opening of the liposome results in releasing of CPPs into the cytoplasm.

CPPs' direct translocation was proposed due to fact that some CPPs cross the cell membrane in an energy-independent process. There are a few different types of structure formed, namely inverted micelles, barrel-staves, and "carpets" (**Figure 5.2**)[83]. In an **inverted micelle**[84], positively charged cell-penetrating peptides firstly interact with negatively charged phospholipids at the cell membrane which induces the formation of a cavity with the CPPs inside. A hydrophobic outer shell makes an inverted micelle possible to move across the membrane's hydrophobic core. It then releases the CPPs (and their binded cargos) into the cytoplasm and becomes part of the cell membrane. The **Barrel-stave model**[85] is a membrane disintegration mechanism, involving three steps. In the first step, peptides bind to the membrane by electrostatic interactions. Next, these peptides insert into the lipid bilayer. Finally, the peptides aggregate into a barrel-like structure in which hydrophilic side chains are pointed toward a central aqueous pore while their hydrophobic parts interact with lipid in the cell membrane. This aqueous pore increases in diameter through the progressive recruitment of additional monomers. The **Carpet model**[85] originated from cellular translocation of antimicrobial peptides. In this mechanism, CPPs bind to the plasma membrane in the orientation parallel to the surface and self-associate with other chains in a carpet-like manner. While its positive side chains stick at the negatively charged phosphate head group, peptide hydrophobic backbones are embedded into the lipid layer of the target cell, perturbing the structural organization

of the membrane. Therefore, peptide that accumulates outside can cross the cell membrane *via* this weakened spot.

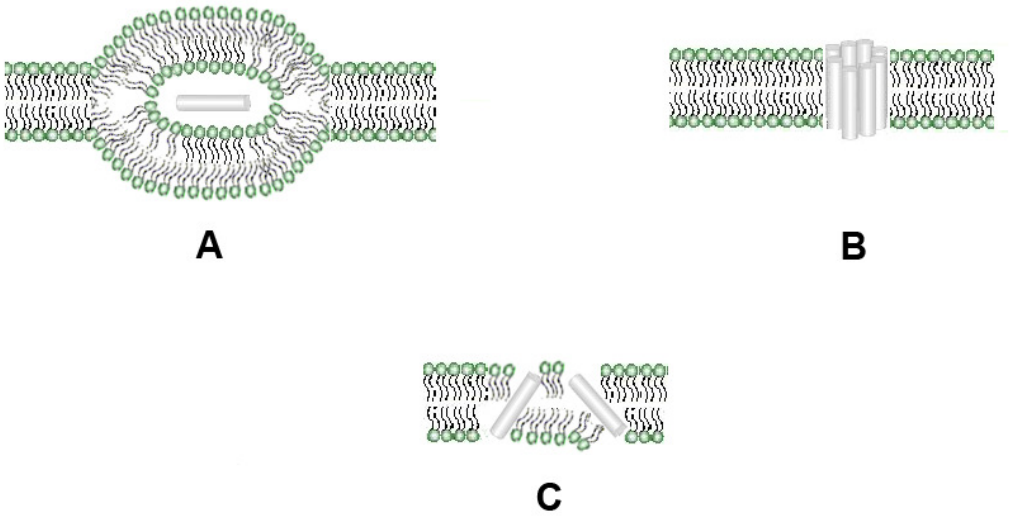


Figure 5.2: peptide-phospholipid arrangements in the inverted micelle (**A**)[84], the barrel-stave (**B**)[85] and the carpet mechanism (**C**)[85].

5.1.3 Applications of cell-penetrating peptides.

5.1.3.1 Drug delivery.

One of the challenges in drug design is the application of a suitable delivery method so that it can be fully effective. For example, cyclosporin A, a cyclic peptide drug used against inflammatory skin diseases, is effective *via* oral administration but with a number of side effects, while direct application on the skin is ineffective due to poor dermal absorption. However, conjugation of a hepta-D-arginine transporter to cyclosporin A (**Figure 5.3**) enhances its skin penetration and makes dermal application possible[86].

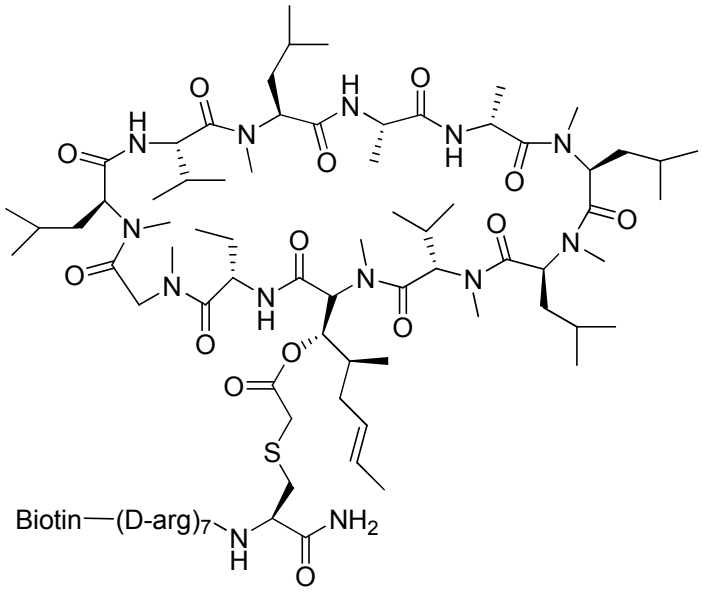


Figure 5.3: hepta-D-arginine-cyclosporin A conjugate.

5.1.3.2 Nucleic acid delivery.

In gene therapy, insertion of highly negatively charged DNA or RNA into a target cell is inefficient. The TAT peptide was found to be able to form complexes with plasmid DNA and therefore carries them into various types of cell *via* endocytosis while siRNAs could be delivered by MPG and polyarginine[87]. Peptide nucleic acids (PNAs), unlike DNA or RNA, contain a neutral peptide backbone but their translocation across plasma membrane is problematic. By conjugating to transportan, a 16-mer PNA could be delivered into HeLa cells[88], while R_9 has also been demonstrated to enable PNA delivery[14].

5.1.3.3 Imaging agent delivery.

Visualization of internal features and the structure of living organisms is important in biological studies. While molecular fluorophores are efficient tool for such studies, their poor cellular uptake is problematic. CPPs could be used to conjugate with dyes to improve their cellular uptake. Applications of selectively

labelling cancer cells has been presented[89]. In addition to binding to a fluorophore, a cationic cell-penetrating peptide sequence was linked with an anionic tail by a cleavable linker. The anionic region acts as an inhibitor of the CPP to cross the cell membrane. In the presented case, an enzyme on the tumor cell, cleaving the anionic region allowed the positively charged cell-penetrating sequence to translocate into the cancer cell, therefore allowing the cancer cells to be identified (**Figure 5.4**).

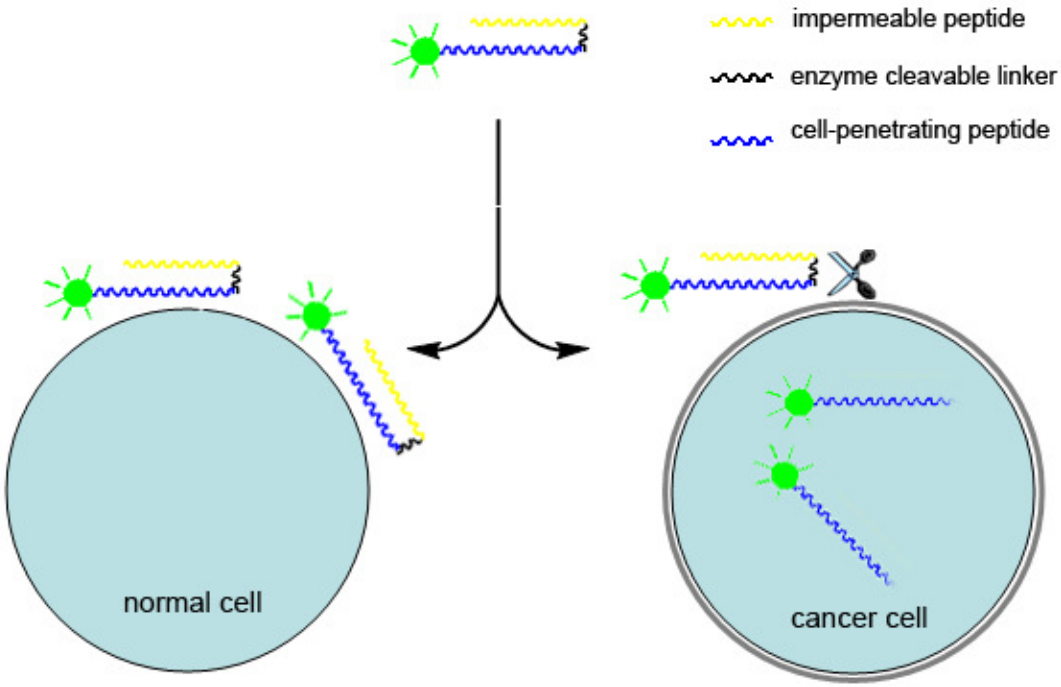


Figure 5.4: Cancer cell identification by protease- triggered CPP release.

5.1.3.4 Protein and macromolecule delivery.

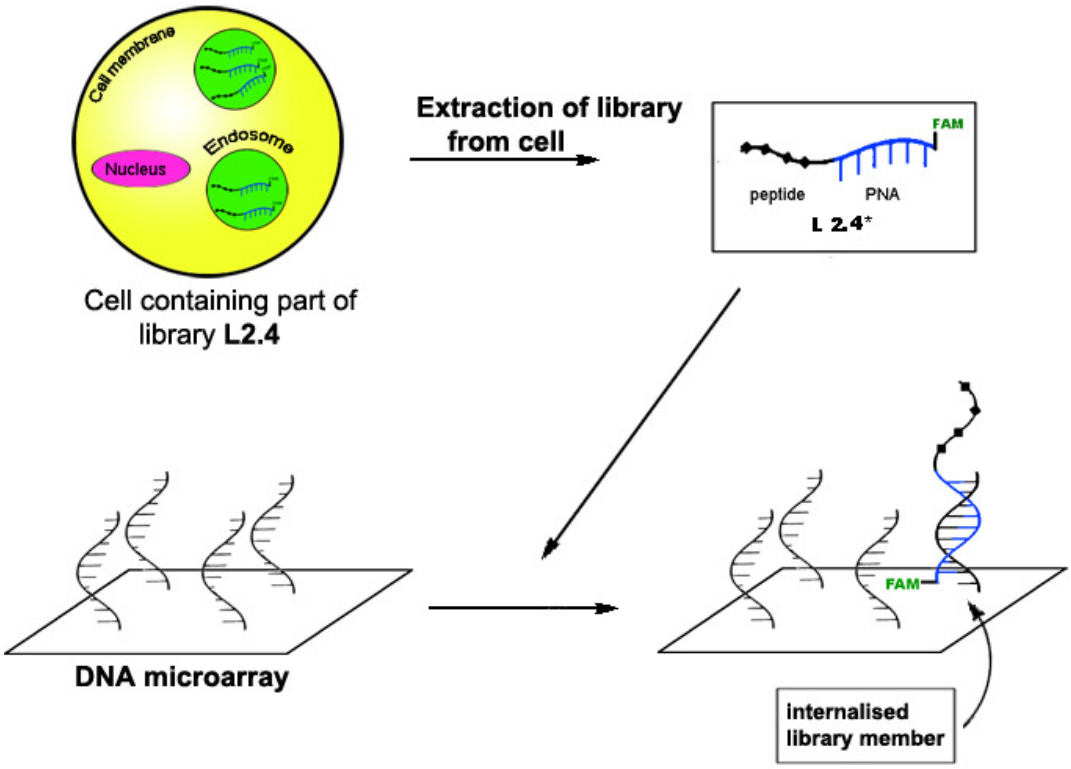
The development of peptide-drugs and therapeutic proteins are often limited by the poor permeability of the cell membrane. However, CPPs can be used as one of the solutions for cellular protein delivery. Pep-1, (KETWWETWWTEWSQP KKKRKV) is an example of a CPP capable of delivering peptides and proteins into cells[90]. Its tryptophan rich region (KETWWETWWTEW) allows hydrophobic

interactions with delivered proteins while the lysine rich motif (KKKRKV) is responsible for membrane crossing. Pep-1 has been used for the delivery of macromolecules including proteins[91,92] peptides[93] and antibodies[94] into various types of cell.

5.2 Results and Discussion.

5.2.1 The screening of a PNA-encoded peptide library - discovering novel cell-penetrating peptides[95].

The procedure for the screening of PNA-encoded peptide library for potential cell-penetrating peptides is summarised in **Scheme 5.1**. A subset of library **L 2.4** (AA^4 = Pro) was added to a mammalian cell culture and incubated for 4 h. The media was then removed and the cells were washed before they were lysed. Members of the subset of library **L 2.4** that were delivered into cells (termed here **L 2.4***) were purified by filter-centrifugation. Purified library members were hybridised onto a DNA microarray containing complementary DNA oligos to all members of library **L 2.4**. The sequences of **L 2.4*** which were taken up by cells were then identified following fluorescence scanning.



Scheme 5.1: Method for the screening of library **L 2.4**[95].

5.2.2 Synthesis of *N*-(6-aminohexyl)glycine tetramer-PNA oligo-conjugate control.

A peptoid consisting of repeating *N*-(6-aminohexyl)glycine units (also known as lysine-like peptoid monomer, Llp) has been shown to be an efficient carrier for cellular delivery[96] and this peptoid was used as a control. A *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate (**5.1, Figure 5.6**) was initially synthesised and delivered into cells. In this compound, *N*-(6-aminohexyl)glycine repeating unit acted as the cellular delivering component, while the PNA 12-mer and FAM were in essence the cargo.

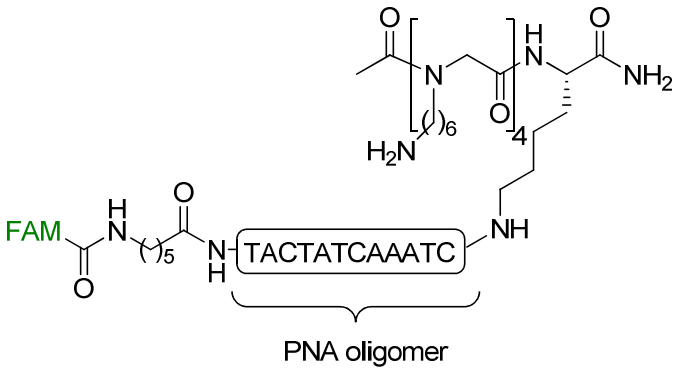
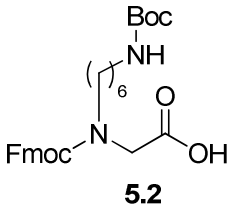


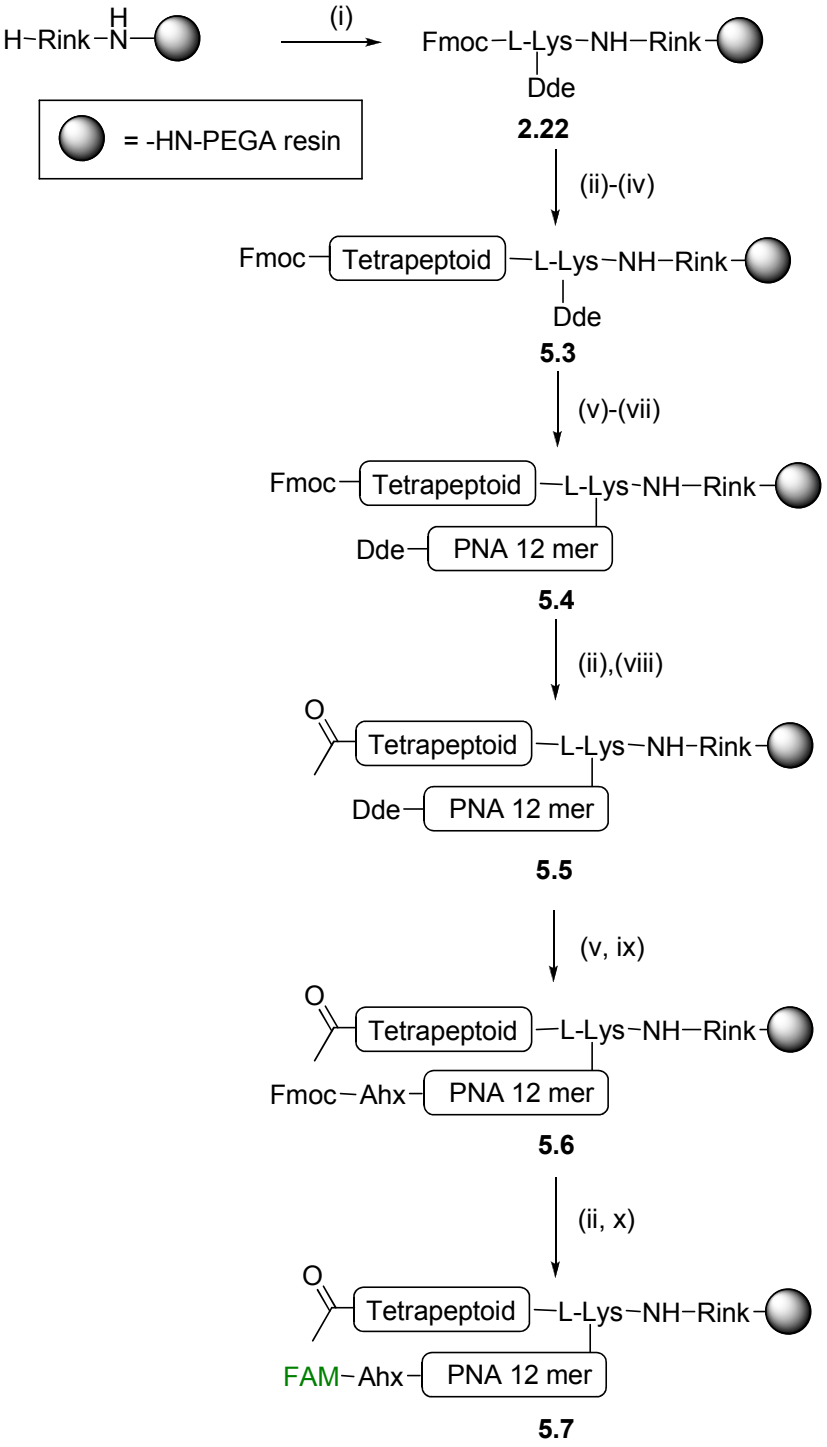
Figure 5.6: *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate (**5.1**).

The synthesis of compound **5.1** is summarised in **Scheme 5.2**. It began by the coupling of Fmoc-Lys(Dde)-OH (**2.19**) to H-Rink-PEGA resin, giving rise to resin **2.22**, which was coupled four times to [(6-*tert*-Butoxycarbonylamino-hexyl)-(9H-fluoren-9-yl methoxycarbonyl)-amino]-acetic acid (**5.2, Figure 5.7**). The resulting resin (**5.3**) was coupled to Dde-PNA monomers until the required length of PNA oligomer (12-mer) was reached (**5.4**). After the removal of the Fmoc group, the free amino group was acetylated with acetic anhydride (**5.5**). An Fmoc-Ahx-OH spacer was coupled to the PNA oligo, followed by the removal of the Fmoc group and labelling with FAM (**5.7**)



5.2

Figure 5.7: [(6-*tert*-Butoxycarbonylamino-hexyl)-(9H-fluoren-9-yl methoxycarbonyl)-amino]-acetic acid.



Scheme 5.2: Synthesis of *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate on solid support; *Reagents and conditions:* (i) **2.19**; (ii) 20% piperidine / DMF, 2 x 10 min; (iii) **5.2**; (iv) repeat (ii),(iii) 4 times; (v) $NH_2OH{\cdot}HCl$/imidazole, 1 h; (vi) Dde-PNA-OH; (viii) repeat (vi),(vii) 12 times; (viii) 1:1 (v/v) Ac_2O:Py, 1.5 h; (ix) Fmoc-Ahx-OH; (x) 5(6)-carboxyfluorescein; (5.0 equiv. amino acids, PNA monomers or dye, 4.8 equiv. PyBOP, 11.0 equiv. NEM/DMF, 60°C, 20 min, microwave irradiation).

After cleavage by a mixture of TFA:TIS:DCM (90:5:5) (**Scheme 2.13**), the *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate (**5.1**, **Figure 5.6**) was obtained by precipitation.

5.2.3 Cellular delivery of *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate.[1]

Solutions of **5.1** (1, 2, 4, 6, 8 and 10 μM) were added to HeLa cells, grown to 90% confluency. After 4 h of incubation, cells were washed and analysed by Flow cytometry. The mean fluorescence of 10,000 cells was recorded and compared with untreated cells. The relationship between fluorescent intensity and concentration of **5.1** is shown in **Figure 5.8**.

[1] Sections 5.2.3, 5.2.4 and 5.2.5 are in collaboration with Nina Svensen

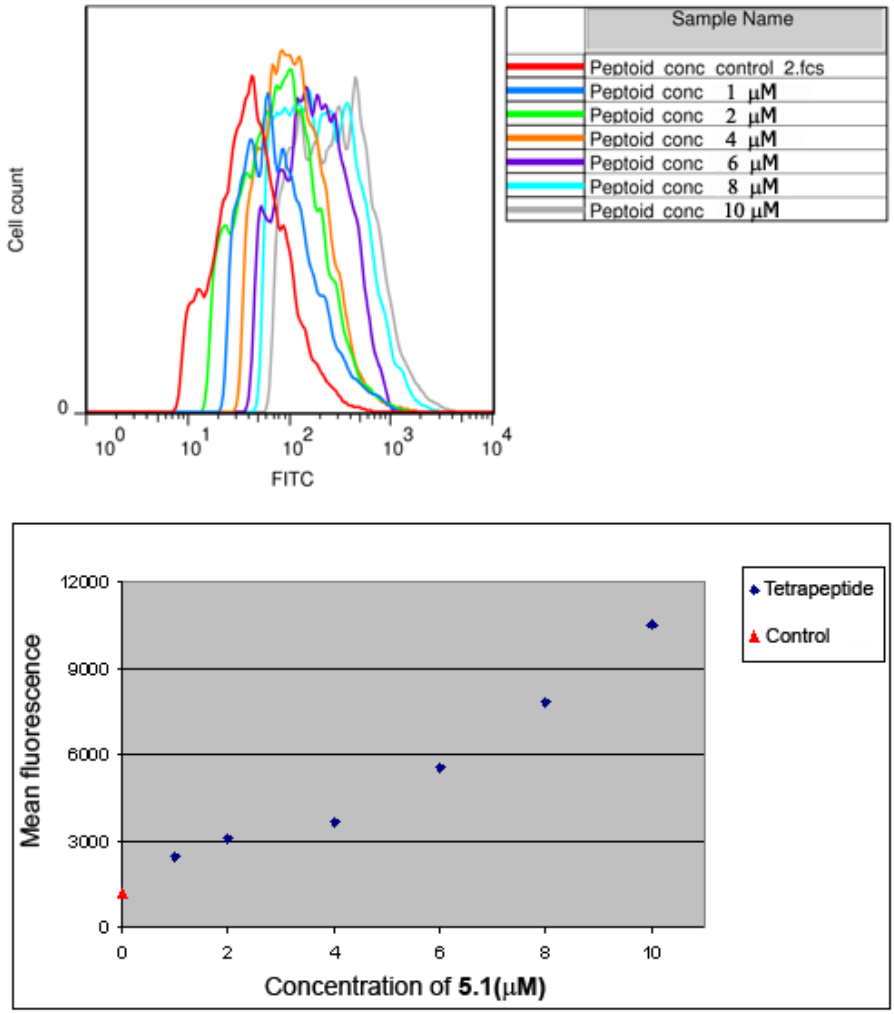


Figure 5.8: Fluorescent intensities of cells after incubation with different concentrations of **5.1**.

Figure 5.8 shows that the mean fluorescence of cells incubated with compound **5.1** correlated with the concentration of **5.1**. It can be concluded that, upon incubation for 4 h, *N*-(6-aminohexyl)glycine tetramer-PNA oligo conjugate (**5.1**) was delivered into HeLa cells.

5.2.4 Cellular delivery of an encoded library into HeLa cells.

A solution of a subset of **L 2.4** [AA^4 = Pro (125 members) (collective concentration 100 µM)] in RPMI media was incubated with HeLa cells grown to 80% confluency for 2 h. After incubation, cells were washed and lysed and the

solution obtained from cells was purified by filter-centrifugation (Millipore, 3,000Da and 10,000Da molecular weight), giving a "cell solution" containing library members that had entered the cells (referred as **L 2.4*** in **Scheme 5.1**). This solution was then hybridised onto a DNA microarray and scanned using the filter set for the detection of FAM. 12 peptide sequences showing the highest fluorescent intensities are summarised in **Figure 5.9**

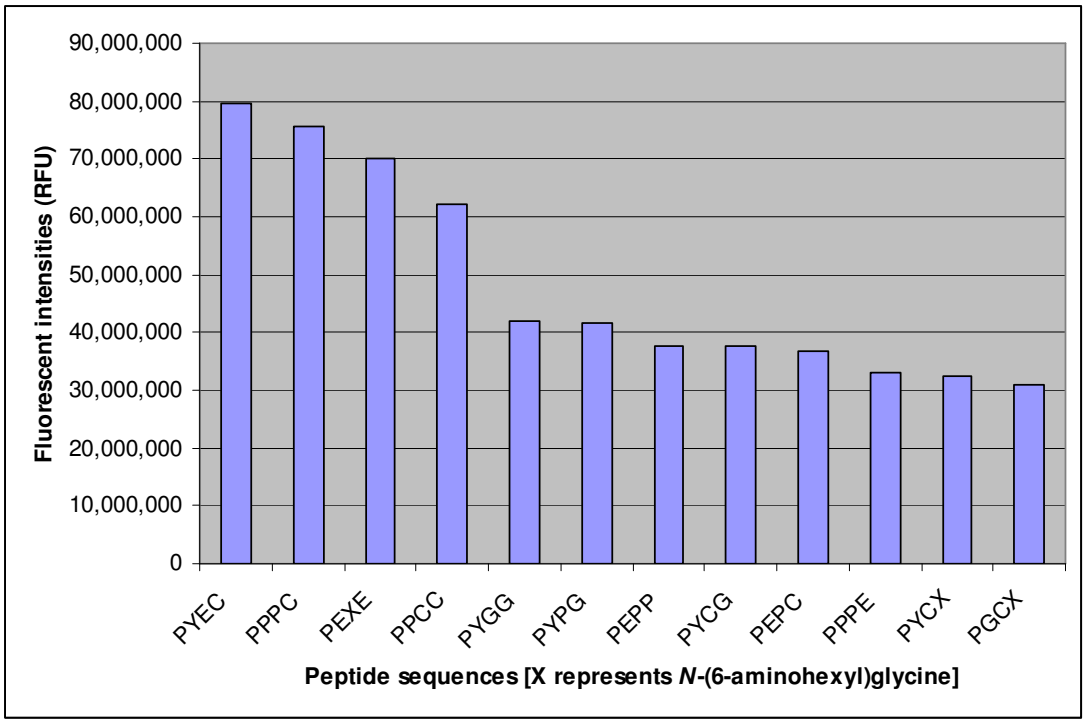


Figure 5.9: Top peptide sequences delivered into HeLa cells using library L **2.4**.

Three of the top four sequences contained Cys at the position AA^1 suggesting that this amino acid is important at this position. Of the top twelve sequences, Tyr was the amino acid most often found at position AA^3 while at position AA^2, no consensus amino acid was found. The consensus "hit" sequences for the cellular delivery of HeLa cells was Pro-Tyr-Yyy-Cys (Yyy represents any amino acid). In addition, Pro-Pro was found four times in the top twelve and is thus a possible peptide sequence delivered to this cell line.

5.2.5 Cellular delivery of an encoded library into B16F10 cells.

Cellular incubation, purification and analysis of the subset of library **L 2.4** [AA^4 = Pro (125 members)] was carried out as described above with B16F10 cell and the 12 peptide sequences having the highest fluorescent intensities are summarised in **Figure 5.10**

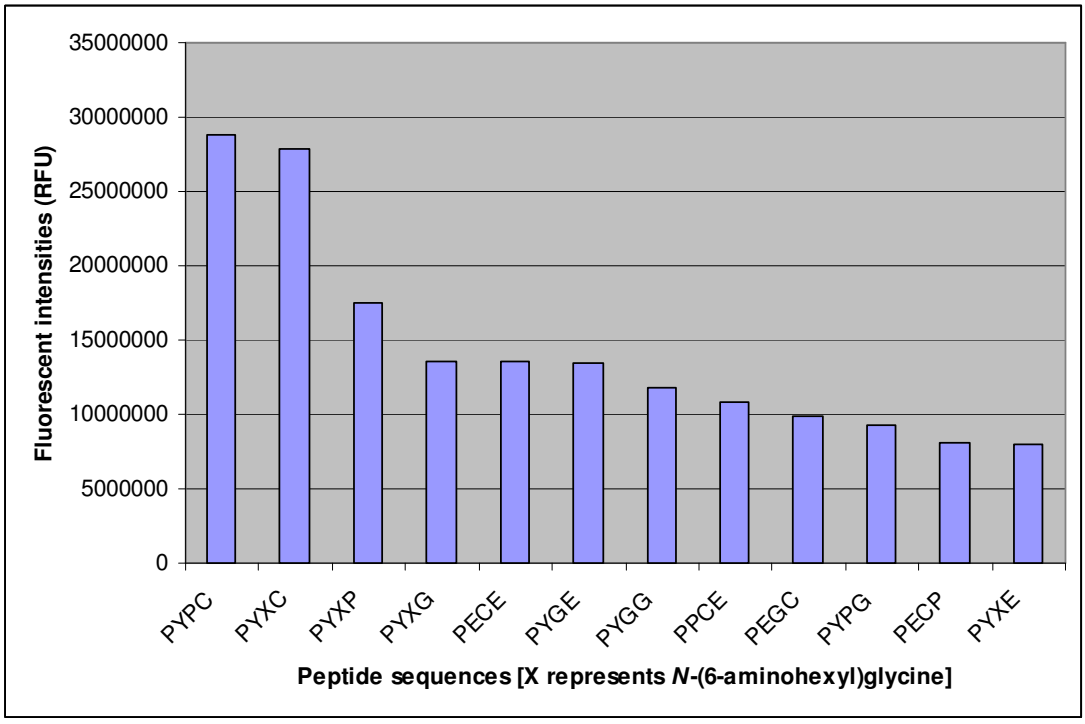


Figure 5.10: Top peptide sequences of **L 2.4** delivered into B16F10 cells.

From the delivery of the subset of library **L 2.4** into B16F10 cells, the top two peptide sequences showed higher efficiency than other sequences and these two sequences have common amino acids, which are Pro, Tyr and Cys at position AA^4, AA^3 and AA^1 respectively. In addition, Tyr was found at position AA^3 in 8 of the top twelve sequences therefore the consensus "hit" sequences for the cellular delivery into B16F10 cells are identical to those for HeLa cells, which is Pro-Tyr-Zzz-Cys where Zzz represents any amino acid. However, in the case of B16F10 cells, the

sequence containing Llp (represented as "X" in Figures) at position AA^2 were preferred.

5.2.6 Synthesis and cellular delivery of the "hits" from the library screen.

A fluorescent-labeled tetrapeptide (FAM-Ahx-Pro-Tyr-Llp-Cys-NH_2, **5.8**) whose sequence was determined from the screening of the libraries was synthesised by solid-phase synthesis as well as a control (FAM-Ahx-Pro-Gly-Ala-Glu-NH_2, **5.9**), which was a negative peptide from the library (**Figure 5.11**). Both peptides were incubated with HeLa and B16F10 cells.

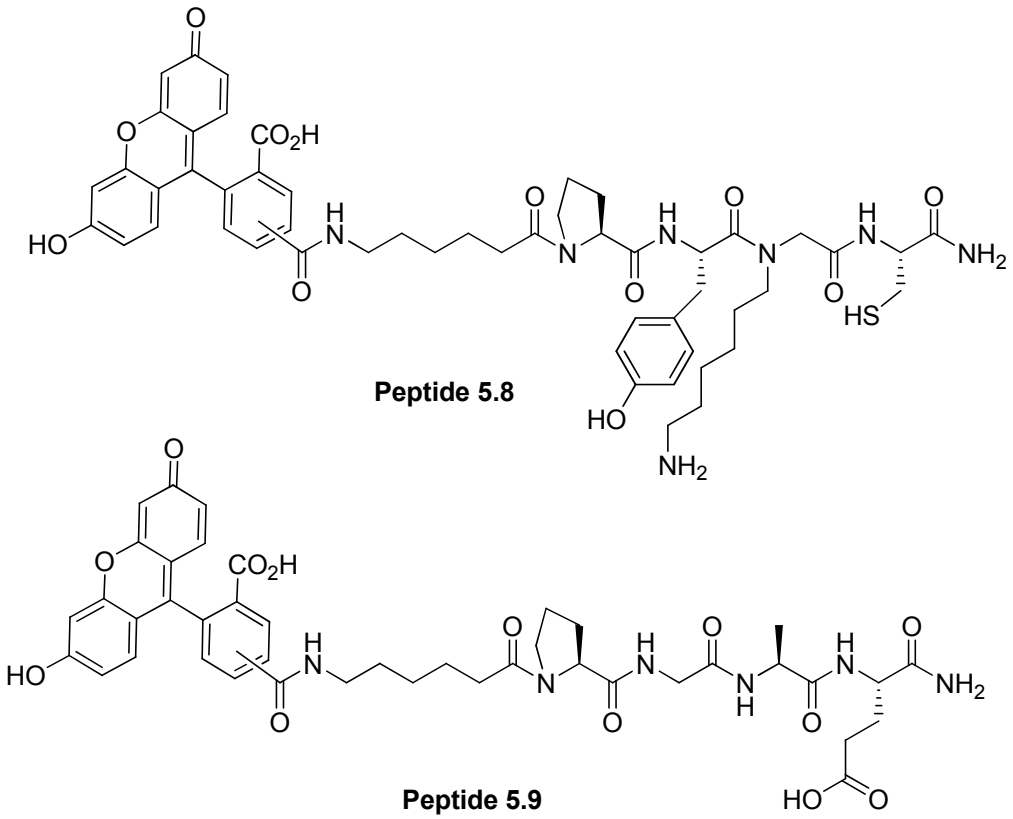


Figure 5.11: Peptide sequences used in evaluation of cellular delivery; consensus "hit" from library screening (**5.8**) and negative control (**5.9**).

*5.2.6.1 Cellular delivery of peptide **5.8** and **5.9** into HeLa cells.*

Solutions of peptide **5.8** and peptide **5.9** were added to the growing media (100 µM). Cells were incubated at 37ºC, 5% CO_2 for 2 h. Cells were detached by trypsination and 0.2% trypan blue added (quenched extracellular fluorescene). Samples were analysed by flow cytometry (using a BD Biosciences FACSDiva and software). Untreated cells were defined as having 0% uptake and the results are shown in **Figure 5.12.**

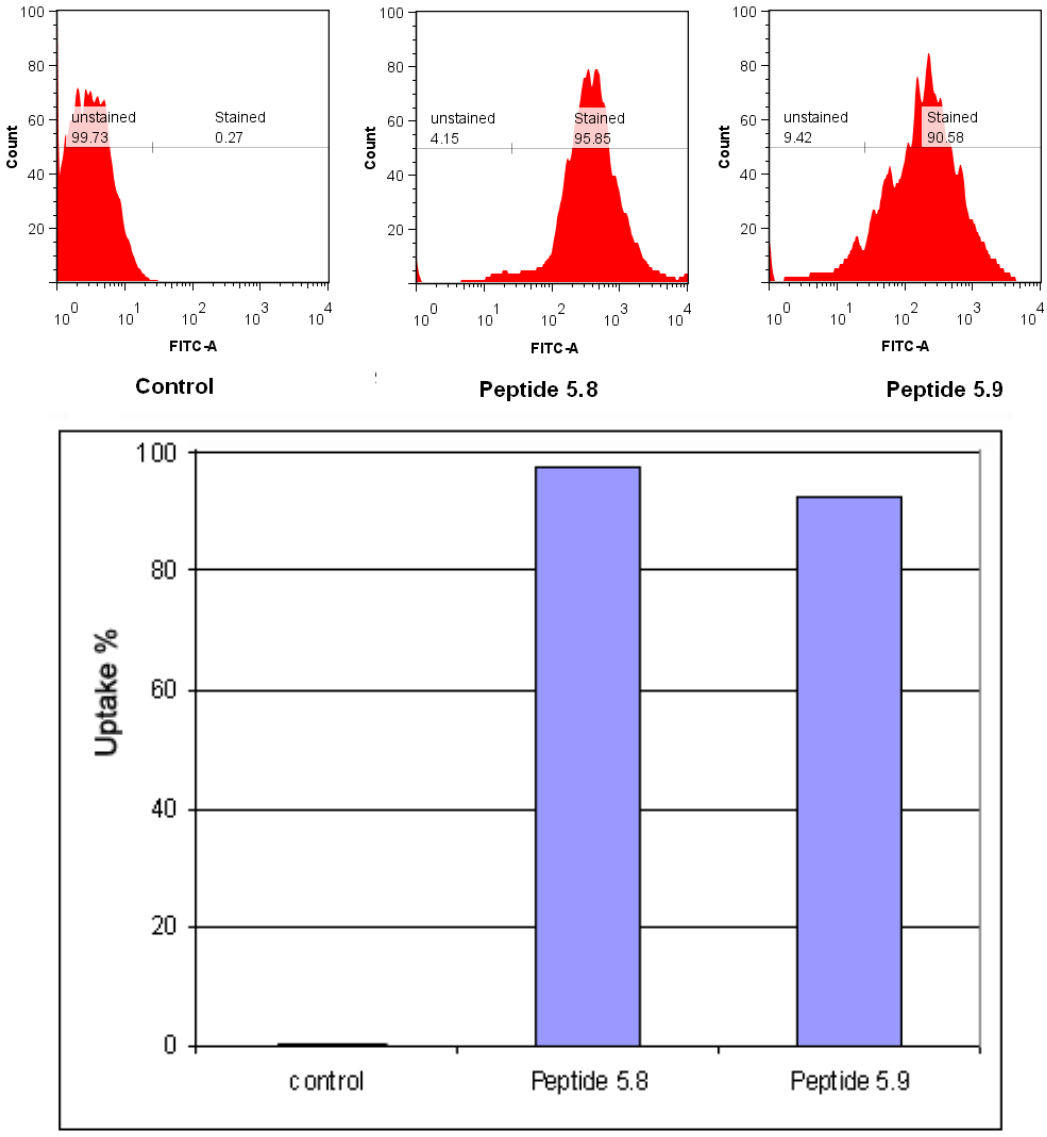


Figure 5.12: Cellular deliveries of peptides **5.8** and **5.9** into HeLa cells.

*5.2.6.2 Cellular delivery of peptide **5.8** and **5.9** into B16F10 cells.*

Cellular delivery into B16F10 cells was carried out as described above and the result is shown in **Figure 5.13**

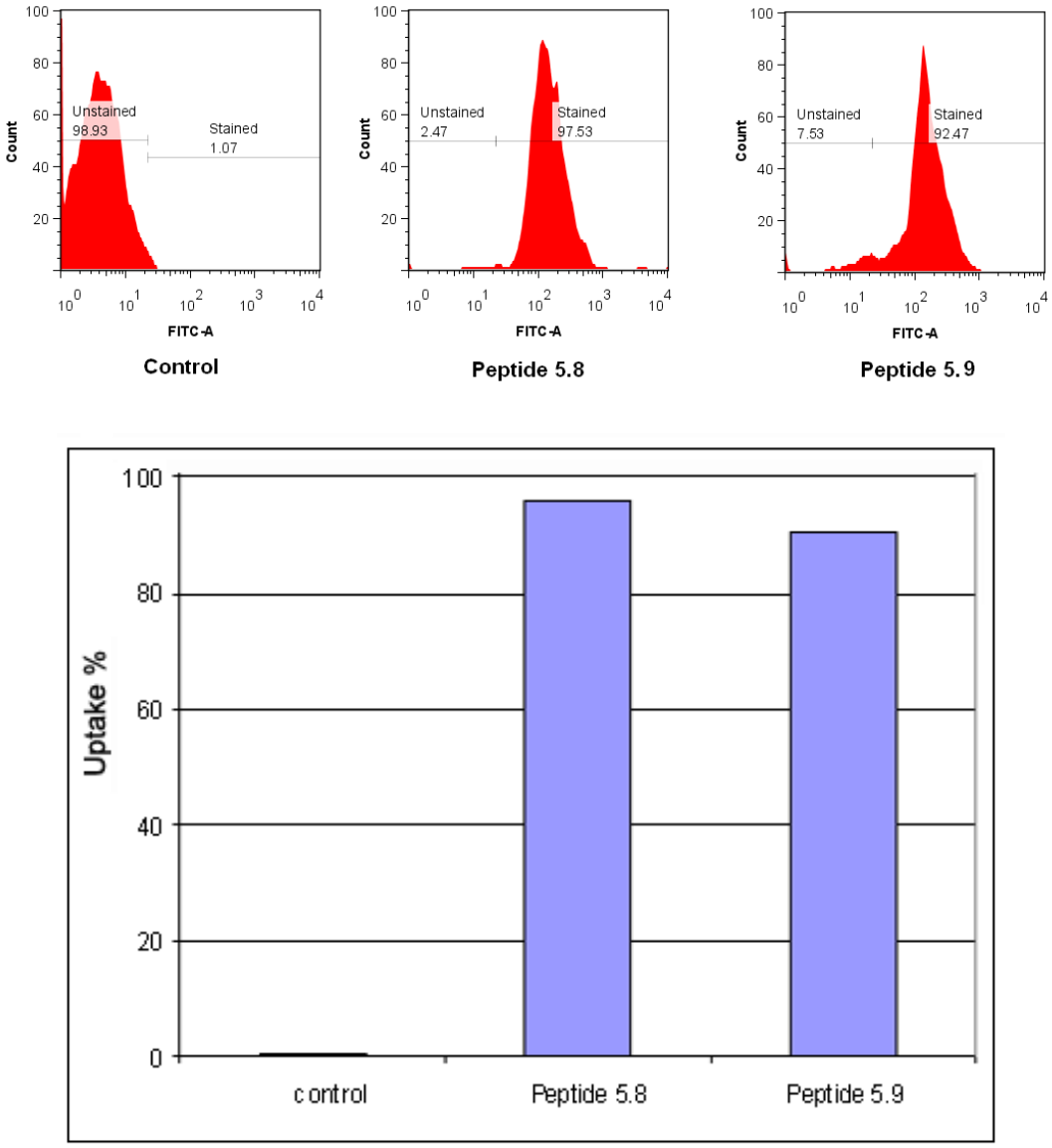


Figure 5.13: Cellular deliveries of peptides **5.8** and **5.9** into B16F10 cells.

While peptide **5.8** was found to be delivered into both cell lines, the negative control (peptide **5.9**) gave the same results. Therefore the experiments were repeated with reduced concentrations of peptides (10 µM). The result of this experiment with B16F10 cells is shown in **Figure 5.14**

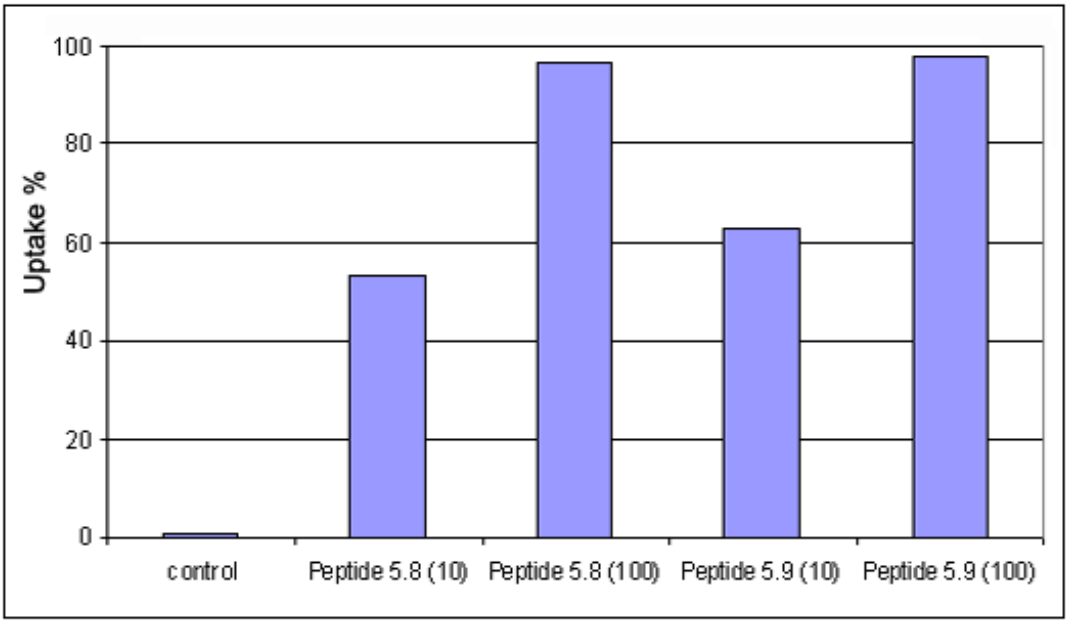


Figure 5.14: Cellular deliveries of peptides and fluorescein at different concentrations into B16F10 cells [numbers in the blankets indicate the concentrations (μM)].

As shown in **Figure 5.14**, the reduced concentration of peptides resulted in a lower uptake %, however, peptide **5.9** (negative control) was still found to be delivered better than peptide **5.8** (the "hit").

*5.2.6.3 Cellular delivery of peptide **5.8** and **5.9** into HeLa cells and comparison with known delivering agents.*

FAM-labeled *N*-(6-aminohexyl)glycine tetramer, FAM-labeled *N*-(6-aminohexyl)glycine nonamer (both known CPPs), peptide **5.8** and peptide **5.9**, were delivered into HeLa cells in order to compare the delivering efficiency. All compounds were incubated with the cells at 10 μM concentration and the uptake % in each case was recorded at 2 h, 4 h and 24 h as shown in **Figure 5.15**

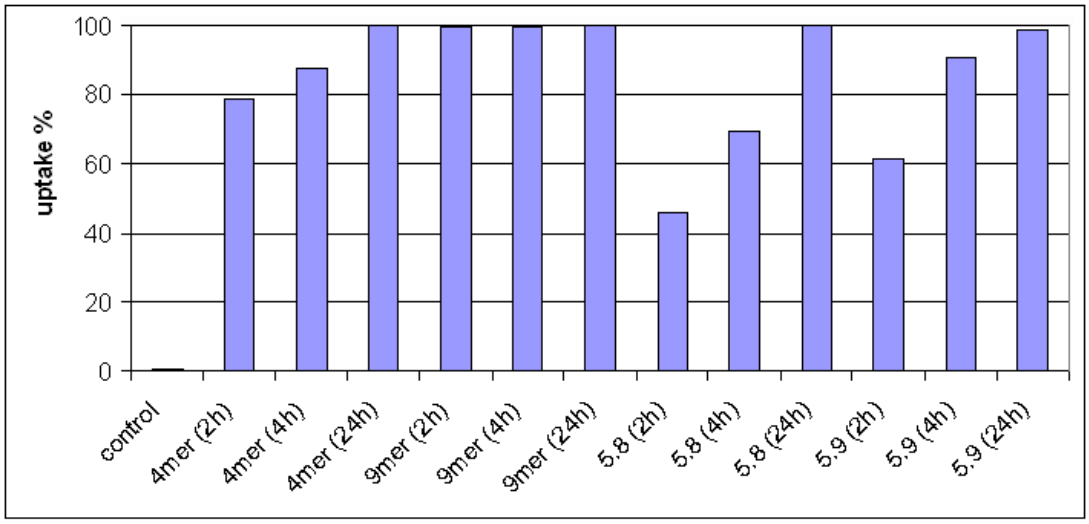


Figure 5.15: Cellular deliveries of *N*-(6-aminohexyl)glycine polymers and peptides into HeLa cells at different times.

The results in **Figure 5.15** show that both peptides (**5.8** and **5.9**) have a cell delivering ability since the cellular uptakes of these peptides were almost complete at 24 h. However, in comparison with *N*-(6-aminohexyl)glycine polymers, peptide **5.8** and peptide **5.9** entered cells at lower rate.

5.3 Conclusions.

In conclusion, a PNA 12-mer as a possible cargo can be delivered into cells by a tetrapeptide CPPs. A PNA-encoded tetrapeptide library was screened for novel CPPs. The "hit" sequences for the delivery into two cell lines (HeLa and B16F10) were found to be identical at three positions (Pro-Tyr-Zzz-Cys where Zzz represents any amino acid) therefore the second amino acid from the carboxyl terminus (AA^2) might be less important in the cellular delivery process.

The sequence of the control (peptide **5.9**) was obtained from a negative member of the library and was expected to be a lower efficiency cellular delivering agent than **5.8**. The rationale for this is that the library used as substrate in the

screening was too small, consisting of only 125 members. It is possible that all members in this library have delivery potential, therefore peptide **5.9**, which had one of the lowest efficiencies of the 125 compound library was still a good delivery agent. In addition, a cargo to be delivered by single peptides (peptides **5.8** and **5.9**) was smaller compared to that delivered by library members (a fluorescein amide versus fluorescein-labelled PNA oligomer respectively), as a result, the delivery by peptide **5.9** was found in high efficiency despite the fact that it showed a poor delivery efficiency in library screening. Therefore, the cargo used in the evaluation of single peptide as delivery agent should have been designed to be identical to that in the library.

Chapter 6: Experimental Section

6.1 General section.

6.1.1 General information.

All solvents and reagents were obtained from commercial suppliers and used without purification, unless otherwise stated. All amino acid derivatives were obtained from GL Biochem (China)

NMR spectra were recorded on a Bruker ARX-250 at 250 MHz for ^{1}H and 62.5 MHz for ^{13}C in the solvents indicated. All chemical shifts (δ) are quoted in ppm using the residual non-deuterated solvent as the internal standard or the ^{13}C natural abundance of the deuterated solvent. Coupling constants (*J*) are reported in Hz.

Infrared (IR) spectra were obtained on a Fourier transform IR Bruker Tensor 27 Spectrometer (FTS) fitted with a Specac single reflection diamond attenuated total reflection (ATR) Golden Gate. All samples were analysed neat and frequencies are reported in cm^{-1} and only frequencies corresponding to significant functional groups are reported.

Aluminium backed silica plates (Merck silica gel 60 F254) were used for thin layer chromatography (TLC) to monitor solution-phase reactions. TLC visualisation was carried out using short wavelength ultraviolet (UV) light (254 nm). Flash column chromatography was performed on silica gel using Keiselgel 60, 230-400 mesh (supplied by Merck).

Microwave reactions were carried out in an Initiator™ single-mode microwave from Biotage (Sweden). HPLC analysis was carried out on an Agilent

Technologies 1100 Chemstation, eluting with (A) H_2O/0.1% TFA and (B) CH_3CN/0.04% TFA at a flow rate of 1 mL/min, with a gradient of 5 to 95% B over 6 min, holding at 95% B for 3 min. The columns used were Discovery C18 from Supelco (50 mm x 4.6 mm, 5 μm), a Luna C18 from Phenomenex (150 mm x 4.6 mm, 5 μm) or a Gemini C18, 110 Å from Phenomenex (100 mm x 4.6 mm, 5 μm). Electrospray ionization mass spectrometry (ESI–MS) analysis was carried out on an Agilent Technologies LC/MSD Series 1100 quadrupole mass spectrometer (QMS) in ESI mode.

Semi-preparative reverse-phase HPLC was performed on an Agilent 1100 system equipped with a Phenomenex Prodigy C18 reverse-phase column (250 x 10 mm, 5 μm) with a flow rate of 2.5 mL/min and eluting with 0.1% TFA in H_2O (A) and 0.042% TFA in acetonitrile (B), with an initial isocratic period of 4 min at 0% B followed by a gradient of 0 to 50% B over 25 min and 50 to 100% B over 10 min, holding at 100% B for 5 min.

6.1.2 General experimental procedures.

6.1.2.1 Qualitative nynhidrin test.[97]

To a few resin beads placed in a small test tube were added 3 drops of reagent A and 1 drop of reagent B. The mixture was heated at 100°C for 5 min. The presence of resin bound free amine was indicated by a blue colour.

Reagent A: Solution 1: Potassium cyanide (65 mg, 1 mmol) was dissolved in H_2O (100 mL). An aliquot of this solution (2 mL) was diluted with freshly distilled pyridine (98 mL). Solution 2: Phenol (40 g, 420 mmol) was dissolved in absolute EtOH (10 mL). Solutions 1 and 2 were mixed together to give reagent A.

Reagent B: Ninhydrin (2.5 g, 14 mmol) was dissolved in absolute EtOH (50 mL).

6.1.2.2 Fmoc deprotection.

To a resin (1 g), pre-swollen in DMF for 10 min, was added a solution of 20% piperidine in DMF (20 mL), and the mixture was shaken for 10 min. The resin was washed with DMF (3 x 25 mL) and the deprotection procedure was repeated. The resin was washed with DMF (3 x 25 mL) and DCM (3 x 25 mL).

6.1.2.3 Dde deprotection.

$NH_2OH{\cdot}HCl$ (6.25 g, 9 mmol) and imidazole (4.59 g, 6.75 mmol) were added to 25 mL of NMP and the mixture was sonicated until complete dissolution. Just before reaction, 5 volumes of this solution were diluted with 1 volume of DMF. The solution was added to the resin and the resulting mixture was shaken for 1 h for PEGA resin and 3 h for PS resin at room temperature. The resin was finally washed with DMF (3 x) and DCM (3 x).

6.1.2.4 Fmoc-amino acid and Dde/Mmt-protected PNA monomers couplings (Method 1).

Fmoc-amino acids or Dde/Mmt-protected PNA monomers (5 eq), PyBOP (4.8 eq) were dissolved in DMF (0.1 M) followed by the addition of NEM (11 eq). The resulting solution was mixed for 10 sec. before adding to the PEGA resin (1 eq) pre-swollen in DMF and the coupling was performed under microwave irradiation at 60ºC for 20 min. Resins were washed with DMF (3 x), DCM (3 x), MeOH (3 x) and DCM (3 x).

6.1.2.5 Fmoc-amino acid and Dde/Mmt-protected PNA monomers couplings (Method 2).[98]

Fmoc-amino acids or Dde/Mmt-protected PNA monomers (4 eq), HOBt (4 eq) and DIC (4 eq) were dissolved in DMF (0.1 M). The resulting solution was mixed for 15 min. before adding to the PEGA resin (1 eq) pre-swollen in DMF and the coupling was performed under microwave irradiation at 60ºC for 20 min. Resins were washed with DMF (3 x), DCM (3 x), MeOH (3 x) and DCM (3 x).

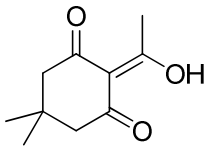
6.1.2.6 Cleavage of libraries from the resin.

Each library was cleaved from the resin (5 g) using a mixture of TFA/TIS/DCM (90/5/5) (1 mL) for 1 h and precipitated with cold diethyl ether. The precipitate was collected by centrifugation and dried *in vacuo* for 2 h to afford library as a solid.

6.2 Experimental to Chapter 2.

6.2.1 Synthesis of PNA building blocks.

2-(1-hydroxyethylidene)-5,5-dimethyl-1,3-cyclohexanedione (2.3).[99]



To a mixture of DCC (75 g, 356 mmol), DMAP (4.4 g, 35.6 mmol) and acetic acid (40.8 mL, 712 mmol) in DMF (580 mL) was added dimedone (50 g, 356 mmol) for 36 h. Precipitated dicyclohexylurea (DCU) was removed by filtration and the

filtrate was concentrate *in vacuo*, taken up in water (300 mL) and extracted with EtOAc (300 mL). The organic layer was washed with 1 M aqueous $KHSO_4$ (3 x 100 mL), 10 % aqueous $NaHCO_3$ (3 x 100 mL) and brine (1 x 100 mL). The organic phase was dried over $MgSO_4$ and concentrated *in vacuo* to afford **2.3** as a mixture of brown oil/yellow solid (58.3 g, 90%).

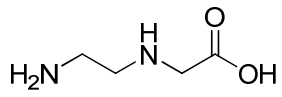
HPLC: t_R = 4.4 min (66% purity, 254 nm).

ESI$^+$/MS: m/z (%) 183.1 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, $CDCl_3$) δ = 2.59 (3H; s, CH_3), 2.52 (2H; s, CH_2), 2.35(2H; s, CH_2), 1.06 (6H, s, $C(CH_3)_2$).

^{13}C NMR (62.5 MHz, $CDCl_3$) δ = 202.4 (C), 197.9 (CO), 195.1(CO), 112.3 (C), 52.4 (CH_2), 46.8 (CH_2), 30.6 (C), 28.5 (CH_3), 28.2(CH_3).

N-(2-aminoethyl)glycine (2.6).[17]



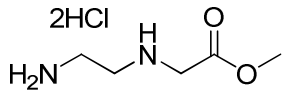
Ethylenediamine **(2.4)** (140 mL, 2.13 mol) was stirred at 4 °C while 2-chloroacetic acid **(2.5)** (20 g, 213 mmol) was added portionwise. The reaction was then stirred for 12 h. Excess **(2.4)** was evaporated *in vacuo* at 60 °C (higher temperature led to decomposition) to give an oil that was triturated with DMSO (500 mL). The resulting white solid was collected by filtration, washed with DMSO (3x 100 mL) and Et_2O (3 x 100 mL) before being dried *in vacuo* at 40 °C for 24 h to give acid **(2.6)** as a white solid (17.6 g, 70%). The product was used without further purification.

ESI$^+$/MS: 192.1 $[M+H]^+$ (28), 214.0 $[M+Na]^+$ (100).

^{1}H NMR (250 MHz, D_2O) δ =3.24 (2H, s, CH_2CO), 3.00 (2H, t, *J* = 6.3 Hz, C**H**$_2$NH), 2.87 (2H, t, *J* 6.3 Hz, NH_2C**H**$_2$).

^{13}C NMR (62.5 MHz, D_2O) δ = 178.2 (CO), 51.6 (CH_2), 46.5 (CH_2), 38.5 (CH_2).

Methyl [(2-aminoethyl)amino]acetate dihydrochloride (2.7).[17]



To a stirred suspension of **2.6** (14.2 g, 120 mmol) in MeOH (450 mL) at 0°C was added $SOCl_2$ (43.5 mL, 600 mmol). The reaction was refluxed for 12 h before cooling to 0°C to give **2.7** as an off-white solid which was collected by filtration, washed with a mixture of MeOH:DCM (1:1) and Et_2O and dried *in vacuo* at 40 °C for 24 h. The product (20.8 g, 85%) was used without further purification.

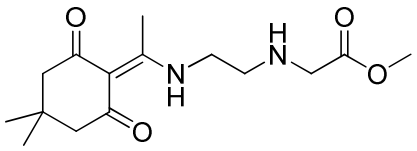
HPLC: t_R = 0.7 min (100% purity, ELSD).

ESI$^+$/MS: m/z (%) 133.1 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, D_2O) δ = 4.11 (2H, s, C**H**$_2$CO), 3.83 (3H, s, C**H**$_3$), 3.54-3.39 (4H, m, 2 x NC**H**$_2$).

^{13}C NMR (62.5 MHz, D_2O) δ = 167.8 (CO), 53.9 (CH_3), 48.0 (CH_2), 44.4 (CH_2), 35.7 (CH_2).

Methyl [(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl) amino]acetate (2.8).[17]



To a stirred solution of diamine salt **(2.7)** (41.6 g, 203 mmol) and DIPEA (71 mL, 406 mmol) in 1:1 DCM:MeOH (v/v) (500 mL) was added Dde-OH **(2.3)** (48.1 g, 264 mmol). The solution was stirred for 16 h before evaporation of the solvent *in vacuo*. The crude material was taken up in EtOAc (400 mL) and extracted with 1 M $KHSO_4$ (3 x 100 mL). The aqueous solution was treated with $NaHCO_3$ until the pH was 9 before extraction with further EtOAc (3 x 100 mL). The organic phase were combined, washed with brine (1 x 100 mL), dried over $MgSO_4$ and concentrate *in vacuo* to give **2.8** (36.0 g, 60%) as a brown oil.

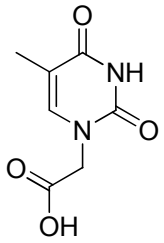
HPLC: t_R = 2.5 min (96% purity, ELSD).

ESI$^+$/MS: m/z (%) 297.2 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, $CDCl_3$) δ = 3.72 (3H, s, OCH_3), 3.47 (4H, m, 2 x NCH_2), 2.93 (2H, t, *J* = 6.1 Hz, NHC**H**$_2$CO), 2.56 (3H, s, CCH_3); 2.35 (4H, s, 2 x CH_2-Dde), 2.03 (1H, br s, NH), 1.01 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (62.5 MHz, $CDCl_3$) δ = 197.3 (CO), 173.1 (C), 172.4 (CO), 107.6 (C), 52.3 (CH_2), 51.5 (CH_3), 49.9 (CH_2), 47.3 (CH_2), 43.1 (CH_2), 29.7 (C), 27.9 (CH_3), 17.8 (CH_3).

(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetic acid (2.10).[10]



To a suspension of thymine **(2.9)** (10.0 g, 79.3 mmol) and K_2CO_3 (10.96 g, 79.3 mmol) in DMF (240 mL) was added methyl bromoacetate (7.35 mL, 79.3 mmol), and the mixture was stirred vigorously overnight under N_2. The reaction mixture was filtrated and evaporated to dryness *in vacuo*. The solid residue was cooled to 0°C, treated with water (80 mL) and 4 M HCl (aqueous, 4 mL), and stirred for 30 min. The precipitate was collected by filtration and washed with water (3 x 40 mL). The precipitate was treated with water (80 mL) and 2 M NaOH (aqueous, 40 mL) and boiled for 10 min. The mixture was cooled to 0°C, treated with 4 M HCl (aqueous, 30 mL), and stirred for 30 min. The white precipitate was collected by filtration, washed with water (3 x 10 mL), and dried over P_2O_5 *in vacuo* for 16 h, giving thymin-1-yl acetic acid **(2.10)** (8.23 g, 56%) as a white solid.

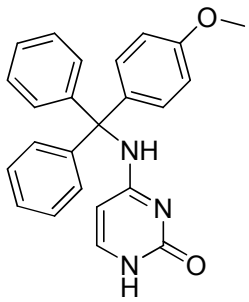
HPLC: t_R = 1.3 min (100% purity, ELSD).

ESI⁻/MS: m/z (%) 183.1 $[M-H]^-$ (100).

^{1}H NMR (250 MHz, d_6-DMSO) δ = 11.34 (1H, br s, NH), 7.49 (1H, s, CH), 4.36 (2H, s, CH_2), 1.75 (3H, s, CH_3).

^{13}C NMR (62.5 MHz, d_6-DMSO) δ = 169.6 (CO), 164.3 (CO), 151.0 (CO), 141.7 (CH), 108.3 (C), 48.4 (CH_2), 11.9 (CH_3).

4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2(1*H*)-pyrimidinone (2.12).[12]



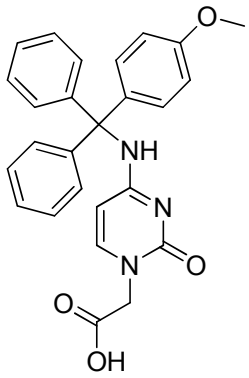
Cytosine **(2.11)** (1.11 g, 10 mmol) and Mmt-Cl (4.63 g, 15mmol) were suspended in pyridine (50 mL) and, after the addition of 4-ethylmorpholine (1.55 g, 10 mmol), this mixture was briefly heated to 45°C and stirred overnight. Water (50 mL) was added to the stirred suspension followed by DCM (30 mL) and the resulting precipitate 1.14 g (30%) was collected by filtration and used in the next reaction without further characterization.

HPLC: t_R = 4.4 min (100% purity, ELSD).

ESI+/MS: m/z (%) 273.0 $[Mmt]^+$ (100), 789.2 $[2M+Na]^+$ (22)

No NMR spectrum was recorded since the named compound **2.12** was insoluble in common deuterated solvents.

(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl)acetic acid (2.13).[17]



NaH (60% in mineral oil, 104 mg, 2.6 mmol) was added to a stirred suspension of 4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2(1*H*)-pyrimidinone **(2.12)** (1.0 g, 2.6 mmol) in DMF (10 mL). After stirring for 1 h, the mixture was cooled to 0°C and methyl bromoacetate (0.3 mL, 3.13 mmol) was added dropwise. The reaction was stirred for 16 h before evaporation of the solvent *in vacuo*. The resulting oil was precipitated with H_2O and collected by filtration, and washed with water to give methyl (4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl) acetate as a white solid which was dried over P_2O_5 *in vacuo.*

Methyl (4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl) acetate (1.4 g, 7.6 mmol) was suspended in 2 N aqueous NaOH (5 mL) and the mixture was stirred for 2 h at reflux. The reaction mixture was cooled to room temperature, washed with DCM (3 x 10 mL) and acidified with 2 N aqueous HCl until pH 2. The precipitate was collected by filtration, washed with water (3 x 10 mL) until the filtrate was neutral and dried *in vacuo* for 16 h to give **2.13** (1.04 g, 91% over 2 steps) as a white solid.

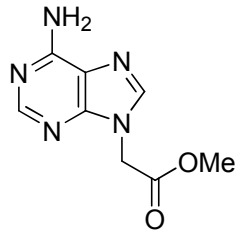
HPLC: t_R = 4.6 min (80% purity, ELSD).

ESI⁻/MS: m/z (%) 454.0 $[M-H]^-$ (100).

^{1}H NMR (250 MHz, *d*$_6$-DMSO) δ = 8.40 (1H, br s, NH), 7.46 (1H, d, J = 6.8 Hz, CH-Cyt), 7.40-7.03 (12H, m, CH-Mmt), 6.84 (2H, d, *J* = 8.5 Hz, CH-Mmt), 6.20 (1H, d, *J* = 6.8 Hz, CH-Cyt), 4.33 (2H, s, CH_2), 3.72 (3H, s, CH_3O-Mmt), 3.61 (3H, s, OCH_3).

^{13}C NMR (62.5 MHz, *d*$_6$-DMSO) δ = 169.1 (CO), 164.0 (C), 157.4 (C), 154.5 (CO), 148.8 (C and CH), 136.7 (C), 129.9 (CH), 128.5 (CH), 127.4 (CH), 126.1 (CH), 112.6 (CH), 95.8 (CH), 69.7 (C), 54.9 (CH_3), 51.9 (CH_3), 49.7 (CH_2).

Methyl (6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl) acetate (2.15).[17]



Adenine **(2.14)** (10.0 g, 74 mmol) was suspended in dry DMF and sodium hydride (60% in mineral oil, 2.14 g, 89 mmol) was added portionwise with stirring. The reaction mixture was stirred for 2 h at room temperature. Subsequently, methyl bromoacetate (13.90 mL, 150 mmol) was added dropwise. After an additional 2 h of stirring, the solvent was removed by evaporation *in vacuo*. The remaining oil was shaken with water (100 mL). The solid was collected by filtration and washed with water (2 x 50 mL) than EtOH (2 x 50 mL) to give 8.60 g (56%) of **2.15** as white crystals.

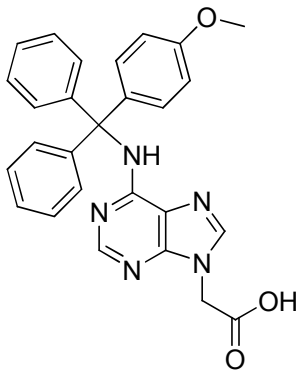
HPLC: t_R = 0.8 min (100% purity, ELSD).

ESI$^+$/MS: m/z (%) 208.1 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, d_6-DMSO) δ = 8.12 (1H, s, CH), 8.11 (1H, s, CH), 7.30 (2H, s, NH_2), 5.08 (2H, s, CH_2), 3.70 (3H, s, CH_3).

^{13}C NMR (62.5 MHz, d_6-DMSO) δ = 168.4 (CO), 155.9 (CNH_2), 152.5 (CH), 149.5 (C), 141.1 (CH), 118.2 (C), 52.3 (CH_3), 43.7 (CH_2).

(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetic acid (2.16).[17]



A mixture of ester **2.15** (3.1 g, 15 mmol), NEM (1.9 mL, 15 mmol) and 4-monomethoxytrityl chloride (6.95 g, 22.5 mmol) in 1:1 pyridine:DCM (100 mL) was heated at 40ºC for 3 h, then 25ºC for 16 h. After evaporation of the solvents *in vacuo*, the residue was re-dissolved in EtOAc (200 mL), and the organic layer were washed with 1 N aqueous $KHSO_4$ (50 mL), 10% aqueous $NaHCO_3$ (50 mL), brine (50 mL), and dried over $MgSO_4$. After evaporation of the solvent *in vacuo*, the residue was washed with petroleum ether (100 mL) to give a yellowish oil (7.42 g). The product was used without further purification.

Methyl (6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl) acetate (3.18 g, 6.62 mmol) was refluxed for 2 h in 1 N aqueous NaOH (100 mL) and stirred at 25°C for 2 h The solution was cooled to 0°C. Following the addition of 1 N aqueous $KHSO_4$ (60 mL), the mixture was stirred at 0°C for 30 min, giving rise to the precipitation of **2.16**. After filtration, the residue was washed with small amount of water (30 mL) until the filtrate was neutral and dried *in vacuo* for 12 h to give **2.16** (3.05 g, 99% over 2 steps) as a white solid.

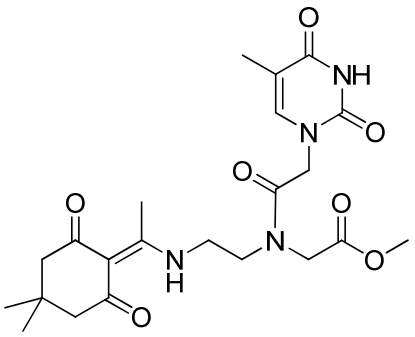
HPLC: t_R = 5.2 min (100% purity, ELSD).

ESI⁻/MS: m/z (%) 464.1 $[M-H]^-$ (100).

^{1}H NMR (250 MHz, d_6-DMSO) δ = 8.20 (1H, s, CH-pur), 7.91 (1H, s, CH-pur), 7.38-7.16 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.9 Hz, CH-Mmt), 4.97 (2H, s, CH_2) 3.71 (3H, s, CH_3O-Mmt).

^{13}C NMR (62.5 MHz, d_6-DMSO) δ = 169.2 (CO), 157.6 (C), 153.4 (C), 151.2 (CH), 148.8 (C), 145.1 (C), 141.9 (CH), 137.0 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.5 (CH), 119.9 (C), 112.9 (CH), 69.8 (C), 54.9 (CH_3), 44.0 (CH_2).

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl) [(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetyl]amino}acetate (2.17 T).[17]



Compound **2.8** (8 g, 27 mmol) was added to a mixture of compound **2.10** (5 g, 27 mmol), DCC (6.6 g, 32 mmol) and HOBt (4.1 g, 27 mmol) in DMF (125 mL). The mixture was stirred for 16 h and DCU was removed by filtration. DMF was evaporated *in vacuo* and the residue was re-dissolved in DCM (300 mL), filtered and the filtrate was washed with 1 N $KHSO_4$ (150 mL), 10% $NaHCO_3$ (150 mL) and evaporated to dryness. The residue was purified by column chromatography, eluting with 5% MeOH/DCM to give **2.17 T** as a white solid (11.5 g, 92%).

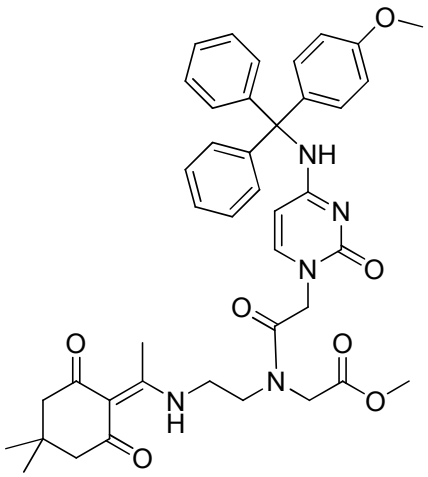
HPLC: t_R = 3.2 min (100% purity, ELSD).

ESI$^+$/MS: m/z (%) 463.3 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: δ = 13.20 and 13.13 (1H, br s, NHDde), 11.32 and 11.30 (1H, br s, NH), 7.33 and 7.31 (1H, s, CH-Thy), 4.67 and 4.51 (2H, s, CH_2COO), 4.41 and 4.11 (2H, s, CH_2CO), 3.72 and 3.62 (3H, s, OCH_3), 3.69 (2H, br s, CH_2N), 3.53 (2H, br s, CH_2N), 2.53 and 2.47 (3H, s, CH_3C-Dde), 2.30 and 2.27 (4H, s, CH_2-Dde), 1.75 (3H, s, CH_3), 0.95 and 0.94 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (62.5 MHz, $CDCl_3$) two rotamers: δ = 174.0 and 173.7 (CO), 169.3 and 169.1 (CO), 168.0 and 167.2 (CO), 164.0 and 163.9 (C), 151.1 and 151.0 (CO), 141.1 and 140.8 (CH), 110.8 and 110.7 (C), 108.4 and 108.3 (C), 53.1 and 52.6 (CH_3), 50.5 (CH_2), 48.4 and 48.2 (CH_2), 47.9 (CH_2), 47.6 (CH_2), 47.3 (CH_2), 41.2 and 40.8 (CH_2), 30.1 (C), 28.2 (CH_3), 18.0 and 17.8 (CH_3), 12.4 (CH_3).

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl)acetyl]amino}acetate (2.17 C).[17]



Compound **2.17 C** was synthesised from **2.13** (5 g, 11.3 mmol) in analogy to **2.17 T** and obtained as a white solid (5.6 g, 69%).

HPLC: t_R = 5.1 min (100% purity, ELSD).

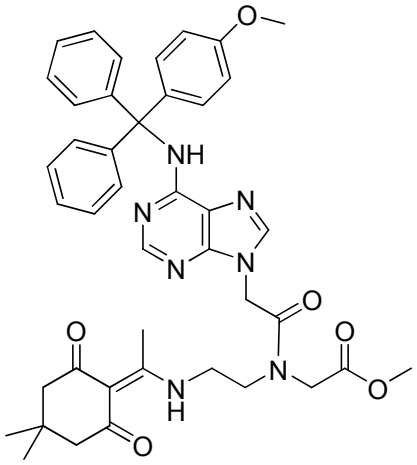
ESI$^+$/MS: m/z (%) 273.1 $[Mmt]^+$ (100), 720.3 $[M+H]^+$ (53).

^{1}H NMR (500 MHz, $CDCl_3$) two rotamers: δ = 13.55 and 13.40 (1H, br s, NH-Dde), 7.25 – 6.94 (13H, m, 12 x CH-Mmt and CH-Cyt), 6.76 and 6.74 (2H, s, CH-Mmt),

4.97 and 4.96 (1H, s, CH-Cyt), 4.51 and 4.34 (2H, s, CH_2COO), 4.33 and 4.03 (2H, s, CH_2CO), 3.73 – 3.64 (7H, m, CH_3O-Mmt and 2 x CH_2N), 3.53 (3H, s, OCH_3) 2.53 and 2.48 (3H, s, CH_3C), 2.26 and 2.25 (4H, s, CH_2-Dde), 0.93 and 0.92 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (125 MHz, $CDCl_3$) two rotamers: δ = 199.1 and 196.7 (CO), 174.1 and 173.8 (C), 169.8 and 169.3 (CO), 168.6 and 167.8 (CO), 165.9 (C), 158.7 (C), 155.8 (CO), 146.2 and 145.9 (CH), 144.1 (C), 135.8 (C), 129.9 (CH), 128.6 (CH), 128.3 (CH), 127.5 (CH), 113.6 (CH), 108.3 and 108.2 (C), 94.7 and 94.6 (CH), 70.6 (C), 60.3 (CH_2), 55.2 (CH_3), 52.8 and 52.5 (CH_3), 51.9 CH_2), 49.0 and 48.7 (CH_2), 48.5 and 48.1 (CH_2), 41.6 and 40.8 (CH_2), 30.0 (C), 28.2 (CH_3), 21.0 (CH_3), 17.9 and 17.7 (CH_3).

Methyl {(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl) [(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetyl] amino}acetate (2.17 A).[17]



Compound **2.17 A** was synthesised from **2.16** (5.5 g, 11.8 mmol) in analogy to **2.17 T** and obtained as a white solid (4.7 g, 53%).

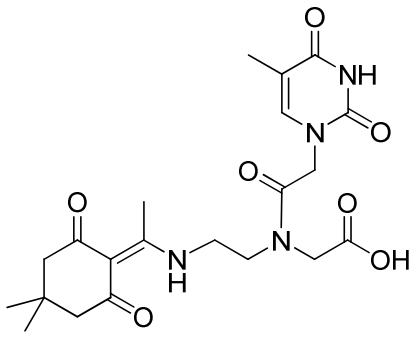
HPLC: t_R = 5.7 min (100% purity, ELSD).

ESI$^+$/MS: m/z (%) 273.0 $[Mmt]^+$ (100), 744.3 $[M+H]^+$ (29).

^{1}H NMR (250 MHz, *d_6*-DMSO) two rotamers: δ = 13.27 and 13.16 (1H, br s, NH-Dde), 8.09 and 8.08 (1H, s, CH-pur), 7.88 (1H, s, CH-pur), 7.33-7.20 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.8 Hz, CH-Mmt), 5.25 and 5.08 (2H, s, CH_2COO), 4.53 and 4.12 (2H, s, CH_2CO), 3.78 (2H, br s, CH_2N), 3.74 and 3.60 (3H, s,OCH_3), 3.72 (3H, s, CH_3O-Mmt), 3.53 (2H, m, CH_2N), 2.56 and 2.46 (3H, s, CH_3C), 2.30 and 2.26 (4H, s, CH_2-Dde), 0.94 and 0.93 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (62.5 MHz, *d_6*-DMSO) two rotamers: δ = 196.5 and 196.4 (CO), 173.2(CO), 169.8 (CO), 169.4 (CO), 167.0 (C), 157.7 (C), 153.4 (C), 151.2(C), 145.2 (C), 142.3 and 142.2 (C), 137.1 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.5 (CH), 119.8 (C), 113.0 (CH), 107.4 and 107.3 (C), 70.0 (C), 54.9 (CH_3), 52.4 (CH_2), 51.8 (CH_3), 48.1 (CH_2), 46.5 (CH_2), 43.8 (CH_2), 29.7 (C), 27.8 (CH_3), 17.4 and 17.1 (CH_3).

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(5-methyl-2,4-dioxo-3,4-dihydro-1(2*H*)-pyrimidinyl)acetyl]amino}acetic acid (PNA-T).[17]



Methyl *N*-[2-(thymin-1-yl)-acetyl]-*N*-{2-[1-(4,4-dimethyl-2,6-dioxo-cyclohexyliden)-ethylamino]-ethyl}-glycinate **(2.17 T)** (2.3 g, 5 mmol) was suspended in 1:1 (v/v) mixture of MeOH and 2 M Cs_2CO_3 (50 mL) and stirred for 1 h. After evaporation of MeOH *in vacuo*, the aqueous phase was acidified with 4 M HCl to pH 1 and the solvents were removed *in vacuo*. The residue was treated with hot 2-propanol (50 mL) and the hot suspension was passed through a filter. The filtrate was collected and 2-propanol was removed *in vacuo*. The residue was sonicated in water (25 mL), the resulting suspension was collected by filtration, giving **PNA-T** as an off-white solid (1.63 g, 73%).

HPLC: t_R = 3.3 min (100% purity, ELSD).

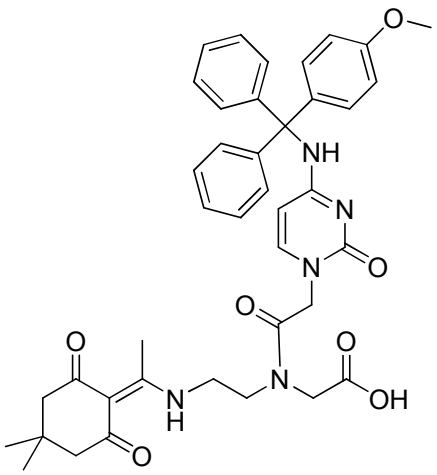
ESI⁻/MS: m/z (%) 895.3 $[2M-H]^-$ (100).

^{1}H NMR (250 MHz, d_6-DMSO) two rotamers: 13.18 and 13.11 (1H, br s, NH-Dde), 11.25 (1H, br s, NH), 7.30 and 7.28 (1H, d, *J* = 1.1 Hz, CH), 4.65 and 4.49 (2H, s, CH_2COO), 4.38 and 4.09 (2H, s, CH_2CO), 3.70 and 3.60 (3H, s, OCH_3), 3.67 (2H, br

s, CH_2N), 3.50 (2H, br s, CH_2N), 2.51 and 2.45 (3H, s, CH_3C-Dde), 2.28 and 2.25 (4H, s, CH_2-Dde), 1.73 (3H, s, CH_3), 0.93 and 0.92 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (125 MHz, $CDCl_3$) two rotamers: δ = 196.9 and 196.7 (CO), 173.5 (CO), 171.1 and 170.7 (CO), 168.4 and 167.7 (CO), 164.7 (C), 151.4 and 151.3 (CO), 142.4 and 142.3 (CH), 108.5 (C), 107.8 and 107.6 (C), 52.8 (CH_2), 49.2 (CH_2), 48.1 and 46.8 (CH_2), 41.3 (CH_2), 30.1 (C), 28.2 (CH_3), 17.7 and 17.6 (CH_3), 12.4 and 12.3 (CH_3).

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(4-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-2-oxo-1(2*H*)-pyrimidinyl)acetyl]amino}acetic acid (PNA-C).[17]



PNA-C was synthesised from **2.17 C** (500 mg, 0.7 mmol) in analogy to **PNA-T** and obtained as a yellowish solid (395 mg, 80%).

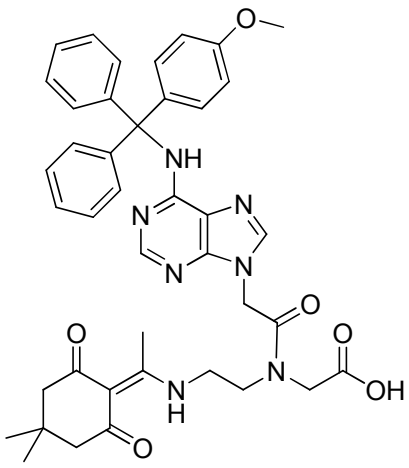
HPLC: t_R = 5.1 min (100% purity, ELSD).

ESI$^-$/MS: m/z (%) 704.2 $[M-H]^-$ (100).

^{1}H NMR (500 MHz, CD_3OD) two rotamers: δ = 7.34-7.24 (11H, m, 10 x CHMmt and CH-Cyt), 7.19 (2H, d, *J* = 8.8 Hz, CH-Mmt), 6.89 (2H, d, *J* = 8.8 Hz, CHMmt), 5.19 and 5.17 (1H, d, *J* = 7.5 Hz, CH-Cyt), 4.70 and 4.57 (2H, s, CH_2COO), 4.29 and 4.12 (2H, s, CH_2CO), 3.79 (3H, s, CH_3O-Mmt), 3.78-3.64 (4H, m, 2 x CH_2N), 2.58 and 2.55 (3H, s, CH_3C), 2.33 and 2.29 (4H, s, 2 x CH_2-Dde), 0.98 and 0.95 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (125 MHz, CD_3OD) two rotamers: δ = 200.1 and 199.8 (CO), 176.0 (C), 172.8 and 172.4(CO), 170.5 and 169.7 (CO), 167.5 and 167.4 (C), 160.5 (C), 148.5 and 148.3 (CH), 145.5 (C), 137.0 (C), 131.7 and 131.4 (CH), 130.2 (CH), 130.0 (CH), 129.5 (CH), 128.7 (CH), 128.6 (CH), 127.5 (CH), 114.7 and 113.8 (CH), 109.2 and 109.1 (C), 96.7 and 96.6 (CH), 72.1(C), 55.8 and 55.7 (CH_3), 53.5 (CH_2), 51.2 (CH_2), 50.7 (CH_2), 49.8 (CH_2), 42.7 and 42.1 (CH_2), 31.1 (C), 28.5 (CH_3), 18.4 and 18.3 (CH_3).

{(2-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}ethyl)[(6-{[(4-methoxyphenyl)(diphenyl)methyl]amino}-9*H*-purin-9-yl)acetyl]amino} acetic acid (PNA-A).[17]



PNA-A was synthesised from **2.17 A** (500 mg, 0.7 mmol) in analogy to **PNA-T** and obtained as a white solid (416 mg, 85%).

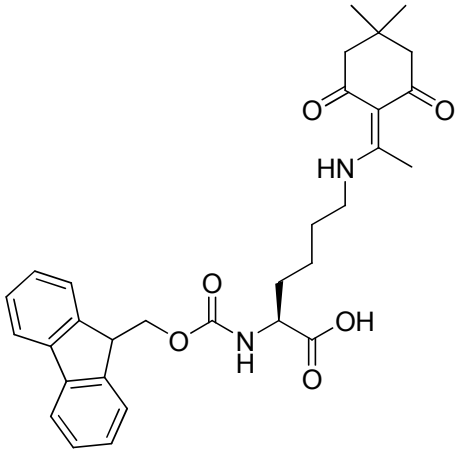
HPLC: t_R = 5.5 min (92% purity, ELSD).

ESI$^-$/MS: m/z (%) 728.2 $[M-H]^-$ (100).

^{1}H NMR (250 MHz, *d_6*-DMSO) two rotamers: 13.25 and 13.15 (1H, br s, NHDde), 8.11 and 8.08 (1H, s, CH-pur), 7.87 (1H, s, CH-pur), 7.33-7.18 (12H, m, CH-Mmt), 6.85 (2H, d, *J* = 8.8 Hz, CH-Mmt), 5.24 and 5.07 (2H, s, CH_2COO), 4.40 and 4.03 (2H, s, CH_2CO), 3.70 (2H, br s, CH_2N), 3.72 (3H, s, CH_3O-Mmt), 3.56 (2H, br s, CH_2N), 2.56 and 2.46 (3H, s, CH_3C), 2.29 and 2.26 (4H, s, CH_2-Dde), 0.93 and 0.92 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (125 MHz, CDCl3) two rotamers: δ = 196.5 (CO), 173.1 (CO), 170.7(CO), 170.2 (CO), 167.6 and 166.8 (C), 157.6 (C), 153.3 (C), 151.1 (CH), 149.0 (C), 145.1 (C), 142.1 (CH), 137.0 (C), 129.8 (CH), 128.4 (CH), 127.6 (CH), 126.5 (CH), 119.7 (C), 112.9 and 112.7 (CH), 107.3 and 107.2 (C), 69.8 (C), 54.9 (CH3), 52.3 (CH_2), 49.1 (CH_2), 47.9 and 46.6 (CH_2), 43.7 (CH_2), 40.9 (CH_2), 29.7 (C), 27.8 and 27.5 (CH_3), 17.3 and 17.2 (CH_3).

(2*R*)-6-{[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl]amino}-2-{[(9*H*fluoren-9-ylmethoxy)carbonyl]amino}hexanoic acid (2.19)[27].



To a stirred suspension of Fmoc-Lys-OH **(2.18)** (5 g, 13.6 mmol) in ethanol (100 mL) was added Dde-OH **(2.3)** (5 g, 27.2 mmol) and TFA (104 µL, 1.36 mmol). The mixture was refluxed for 60 h and concentrated *in vacuo*. The orange residue was dissolved in EtOAc (200 mL), washed with 1 M aqueous $KHSO_4$ (2 x 100 mL), brine (1 x 100 mL), dried over $MgSO_4$ and the solvent was evaporated *in vacuo*. The resulting yellow oil was triturated with hexane to give a yellowish solid that was recrystallised from ethyl acetate/hexane to afford Fmoc-Lys(Dde)-OH **(2.19)** as an off-white solid (5 g, 69%).

HPLC: t_R = 5.1 min (90% purity, ELSD).

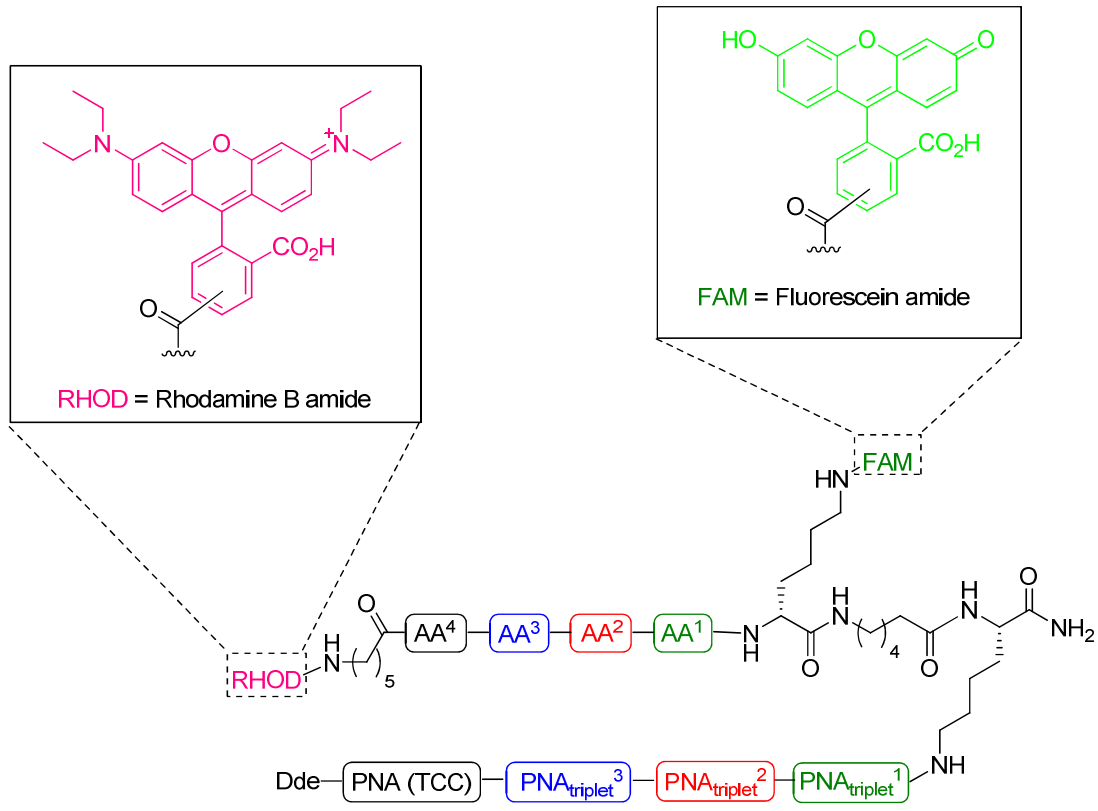
ESI$^+$/MS: m/z (%) 533.2 $[M+H]^+$ (100).

^{1}H NMR (250 MHz, d_6-DMSO): δ = 13.34 (1H, br s, COOH), 12.72 (1H, br s, NHDde), 7.97 (2H, d, *J* = 7.4 Hz, CH-Fmoc), 7.81 (2H, d, *J* = 7.2 Hz, CH-Fmoc), 7.50 (2H, t, *J* = 7.4 Hz, CH-Fmoc), 7.42 (2H, d, *J* = 7.2 Hz, CH-Fmoc), 5.84 (1H, br s, NH), 4.41-4.38 (1H, m, CH), 4.33 (2H, d, *J* = 8.2 Hz, CH2-Fmoc), 4.09-3.95 (1H, m, CH-Fmoc), 3.57-3.49 (2H, m, CH_2-NH), 2.55 (3H, s, CH_3-Dde), 2.32 (4H, s,

CH_2-Dde), 1.91-1.72 (2H, m, CH_2-Lys), 1.71-1.56 (2H, m, CH_2-Lys), 1.55-1.37 (2H, m, CH_2-Lys), 1.00 (6H, s, $(CH_3)_2$-Dde).

^{13}C NMR (62.5 MHz, d_6-DMSO): δ = 196.3 (CO), 173.8 (CO), 172.7 (C=CNH), 156.1 (COO), 143.7 (C), 140.6 (C), 127.5 (CH), 127.0 (CH), 125.2 (CH), 120.0 (CH), 106.8 (C), 65.5 (CH_2), 53.6 (CH), 52.2 (2 x CH_2), 46.6 (CH), 42.4 (CH_2NH), 30.3 (CH_2), 29.6 (CH_2), 28.0 ($C(CH_3)_2$), 27.8 ($C(CH_3)_2$), 23.0 (CH_2), 17.3 (CH_3).

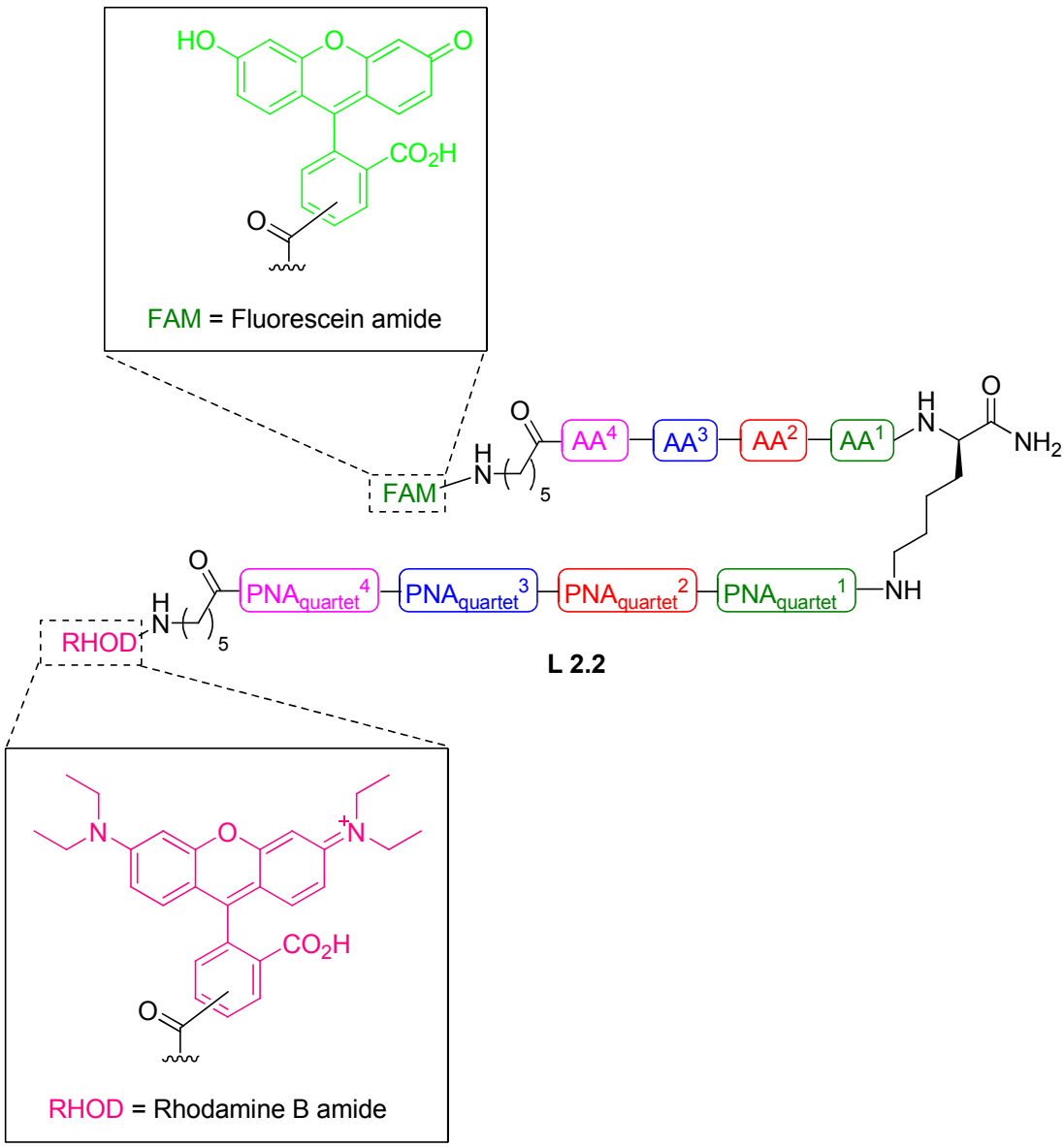
6.2.2 Synthesis PNA-enocoded FRET based D-amino acid containing peptide library (L 2.1).



Deprotection, coupling (method 1) and cleavage protocols are described in section 6.1.2. Fmoc-Rink-amide linker (53 mg, 100 µmol, 5 eq) was first attached to

PEGA resin (337 mg, 20 μmol) followed by Fmoc deprotection and stepwise couplings with Fmoc-Lys(Dde)-OH **2.19** (53 mg, 100 μmol, 5 eq), Fmoc-Ahx-OH (35 mg, 100 μmol, 5 eq), and Fmoc-Lys(Alloc)-OH **2.21** (45 mg, 100 μmol, 5 eq). The synthesis was then carried out as follows: [1] split resin **2.24** into 5 pools (4 μmol resin loading each pool) [2] Fmoc deprotection; [3] Fmoc-amino acid couplings (20 μmol per pool, 5 eq; method 1, section 6.1.2.4); [4] Dde deprotection; [5] Dde/Mmt-protected PNA monomer couplings (20 μmol per pool, 5 eq; method 1, section 6.1.2.4); repeat [4] and [5] twice (amino acids and Dde/Mmt-protected PNA monomers used in steps [3] and [5] were according to **Table 2.3**); [6] the five resin pools were then mixed, repeat steps from [1] to [6] twice. Steps [1] to [5] were repeated using "TCC" as PNA code (*N* to *C*), giving raise to **2.26** (in 5 pools). For each pool, Fmoc-Ahx-OH (7.0 mg, 20 μmol, 5 eq) and 5(6)-carboxyrhodamine B (18.6 mg, 20 μmol, 5 eq) were respectively coupled followed by Alloc-deprotection and coupling with 5(6)-carboxyfluorescein (7.5 mg, 20 μmol, 5 eq). Before resin cleavage, the sub-library was treated with 20% piperidine in DMF (6 mL, 2 cycles of 10 min)[97], washed with DMF (2 x 5 mL) and DCM (2 x 5 mL). Final cleavage from the resin was carried out using TFA/TIS/DCM (90/5/5) (1 mL) for 1 h to afford the sub-library of **L 2.1** as a magenta solid after precipitation with cold Et_2O.

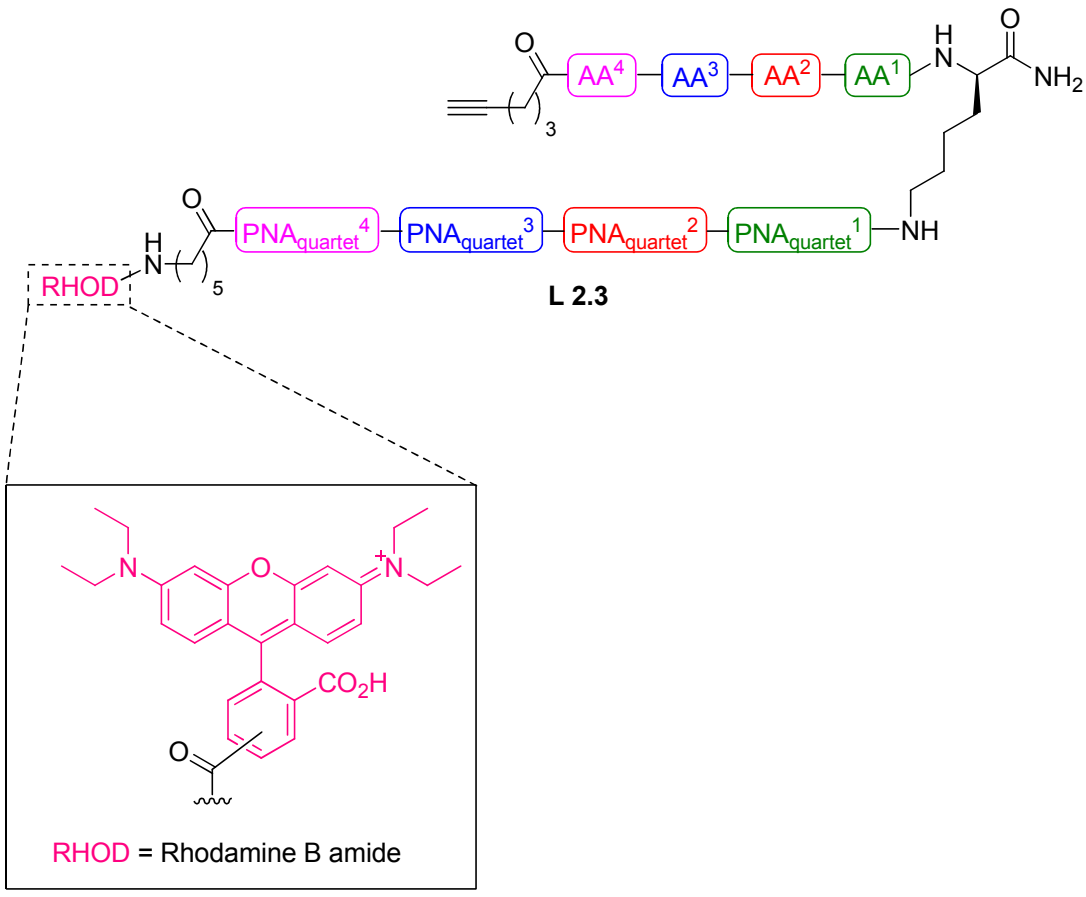
6.2.3 Synthesis of a 1296 member D-amino acid containing peptide library (L 2.2).



Deprotection, coupling (method 2) and cleavage protocols are described in section 6.1.2. Fmoc-Rink-amide linker (77 mg, 142 µmol, 4 eq) was first attached to the PEGA resin (600 mg, 36 µmol) followed by Fmoc deprotection and stepwise couplings with Fmoc-D-Lys(Dde)-OH (**2.31**). The synthesis was then carried out as follows: [1] split resin **2.32** into 6 pools (6 µmol resin loading each pool); [2] Fmoc

deprotection; [3] Fmoc-amino acid couplings (24 µmol per pool, 4 eq; method 2, section 6.1.2.4); [4] Dde deprotection; [5] Dde/Mmt-protected PNA monomer couplings (24 µmol per pool, 4 eq; method 2, section 6.1.2.4); repeat [4] and [5] three times (amino acids and Dde/Mmt-protected PNA monomers used in steps [3] and [5] were according to **Table 2.4**); [6] the five resin pools were then mixed, repeating steps from [1] to [6] three times, giving raise to resin **2.34**. To one fourth of resin **2.34** (9 µmol), Fmoc-Ahx-OH (12.7 mg, 36 µmol, 4 eq) and 5(6)-carboxyfluorescein (13.5 mg, 36 µmol, 4 eq) were respectively coupled followed by Dde-deprotection and coupling with 5(6)-carboxyrhodamine B (17.5 mg, 36 µmol, 4 eq). The library was cleaved using the same condition as for **L 2.1**.

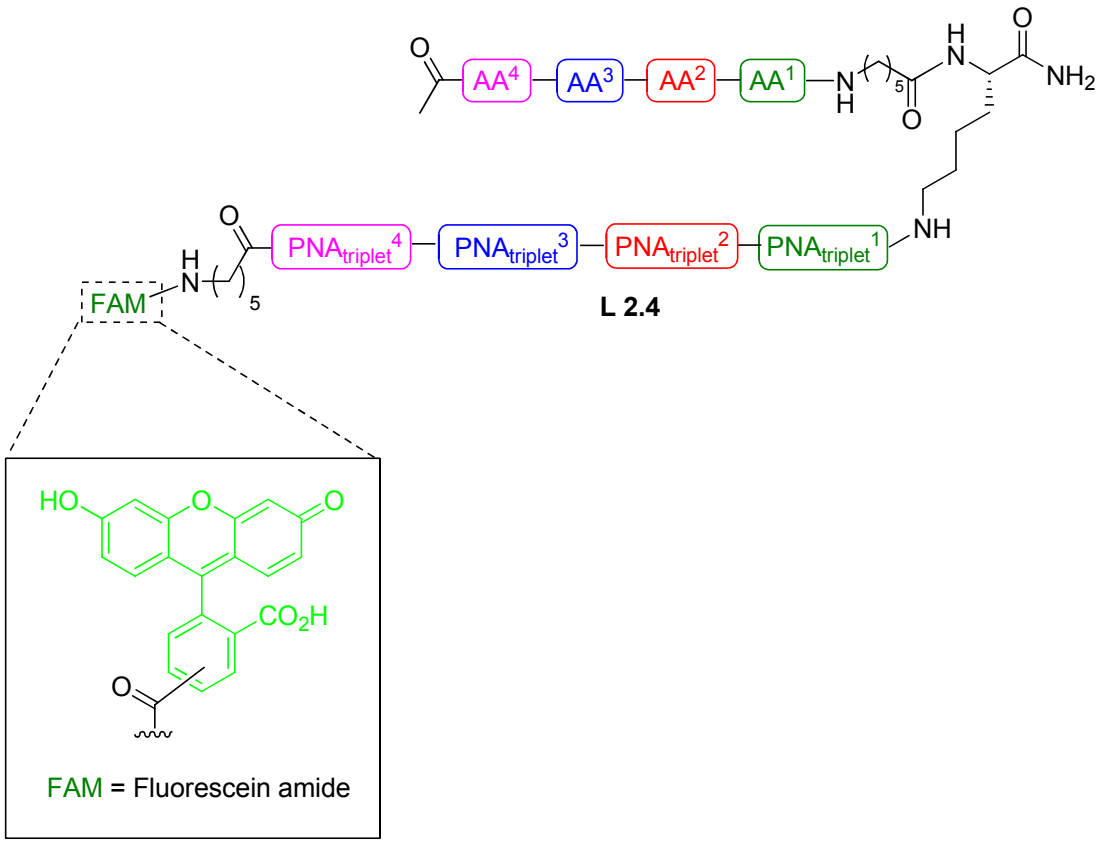
6.2.4 Synthesis of a 1296 membered alkyne peptide library (L 2.3).



L 2.3

Resin **2.33** (18 µmol) was coupled (method 2) to 5-hexynoic acid (7.9 µL, 72 µmol, 4 eq) followed by Fmoc-deprotection. After Dde-deprotection, Fmoc-Ahx-OH (25.4 mg, 72 µmol, 4 eq) and 5(6)-carboxyrhodamine B (35.2 mg, 72 µmol, 4 eq) were respectively coupled. The library was cleaved using the same conditions as for **L 2.1** (see section 6.1.2.6).

6.2.5 Synthesis of a 625 member PNA-encoded peptide library of potential cell-penetrating peptides (L 2.4).



L 2.4

Deprotection, coupling (method 2) and cleavage protocols are described in section 6.1.2. Fmoc-Rink-amide linker (53 mg, 100 µmol, 5 eq) was first attached to PEGA resin (337 mg, 20 µmol) followed by Fmoc deprotection and coupling with Fmoc-Lys(Dde)-OH **2.19** (53 mg, 100 µmol, 5 eq) and Fmoc-Ahx-OH (35 mg, 100 µmol, 5 eq). The synthesis was then carried out as follows: [1] split resin **2.23** into 5

pools (4 μmol resin loading each pool); [2] Fmoc deprotection; [3] Fmoc-amino acid couplings (20 μmol per pool, 5 eq; method 2, section 6.1.2.4); [4] Dde deprotection; [5] Dde/Mmt-protected PNA monomer couplings (20 μmol per pool, 5 eq; method 2, section 6.1.2.4); repeat [4] and [5] twice (amino acids and Dde/Mmt-protected PNA monomers used in steps [3] and [5] (see **Table 2.5**); [6] the five resin pools were then mixed, repeating steps from [1] to [6] twice. Steps [1] to [5] were repeated and the resin was left in 5 portions. Each portion was treated with 1:1 Ac_2O:Py (2 mL) for 1.5 h following Fmoc deprotection. After the Dde deprotection, Fmoc-Ahx-OH (7.0 mg, 20 μmol, 5 eq) and 5(6)-carboxyfluorescein (7.5 mg, 20 μmol, 5 eq) were respectively coupled. Before cleavage, the sub-library was treated with 20% piperidine in DMF (6 mL, 2 cycles of 10 min)[100], washed with DMF (2 x 5 mL) and DCM (2 x 5 mL). Final cleavage from the resin was carried out using TFA/TIS/DCM (90/5/5) (1 mL) for 1 h to afford the sub-library of **L 2.4** as a yellowish solid after precipitation with cold Et_2O.

6.3 Experimental to Chapter 3.

6.3.1 Screening of library L 2.1 as potential substrate for chymopapain.

PBS buffer: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4.

Enzyme Buffer: 50 mM sodium acetate, 2 mM cysteine, 0.1 mM EDTA, pH 6.2.

A solution of library **L 2.1** (16 μL, 3.4 nmol) in DMSO was diluted into 59 μL of PBS buffer and heated at 90°C for 20 min. The solution was cooled to room temperature, 20 μL of 3 μM chymopapain was added and the mixture was incubated at 37°C for 18 h. One volume of this solution was diluted with one volume of

GenHyb buffer (Genetix) and denatured by heating at 90°C for 20 min. 100 µL of the solution was added onto a DNA microarray (see **3.4.1**) at 60°C. The temperature was then lowered to 25°C over a period of 20 h. After hybridisation, the chips were washed with 4x SSC (20 mL); washed twice with 2x SSC, 0.1% SDS (20 mL) at 25°C for 5 min; 0.2x SSC for 1 min; 0.1x SSC for 1 min and spin-dried by centrifugation (1000 rpm for 10 min).

6.3.2 Screening of library L 2.2 as a potential substrate for chymopapain.

Library **L 2.2** was screened as described in **6.3.1** except that the reaction mixture was hybridised onto a 22,500 customised DNA chip from Oxford Gene Technology (Oxford, UK) and after the hybridisation, the chips were washed twice (20 mL) for 10 min at 30°C with the following buffer (100 mM NaCl, 10 mM citric acid, 0.7% (w/v) N-lauroylsarcosine sodium salt, 0.1 mM EGTA, pH 7.5), rinsed with distilled water (20 mL) and spin-dried by centrifugation (1000 rpm for 10 min).

6.3.3 Screening of library L 2.2 as a potential substrate for Gumby.

Enzyme Buffer: 20 mM HEPES, 100mM NaCl, 2 mM DTT, pH 7.5

Library **L 2.2** (60 µL, 600 pmol) in PBS buffer was screened against Gumby (2 µL, 11.4 pmol) as described in **6.3.1** except that the reaction mixture was hybridised onto a 22,500 customised DNA chip from Oxford Gene Technology (Oxford, UK). After the hybridisation, the chips were washed as described in **6.3.2**.

6.4 Experimental to Chapter 4.

6.4.1 Screening of library L 2.3 as potential substrates for a self promoting Huisgen 1,3-dipolar cycloaddition.

An aqueous mixture (25 μL) of **L 2.3** (20 μM) and compound **4.2** (20 μM) was incubated at 37°C for 18 h. Excess **4.2** was removed by a centrifugal filter device (Millipore) and the reaction mixture was concentrated by freeze drying. The concentrated library was solubilised in 50 μL of water and diluted with 50 μL of 2x GexHyb buffer (Genetix) and hybridised onto a 22,500 customised DNA microarray as described in **6.3.2**.

6.5 Experimental to Chapter 5.

6.5.1 Screening of library L 2.4 for the discovery of cell-penetrating peptides.

The screening of library **L 2.4** was performed by Nina Svensen as a part of her PhD thesis[101].

6.5.2 Cellular delivery of "hit" from CPP library.

Cells were suspended to the appropriate cell density in fresh growth media - Roswell Park's Memorial Institute (RPMI, sigma aldrich) for HeLa cells and Dulbecco's Modified Eagle Medium (DMEM, sigma aldrich) for B16F10 cells - before seeding onto polystyrene well-plates (Nunc). Cells were incubated (37 ºC/5% CO_2) for 24 hours to allow adhesion. The peptides were added into the growing media to obtain the desired final concentration. Cells were incubated in the presence of peptide **5.2** and peptide **5.3** (37 ºC/5% CO_2).

6.5.3 Flow cytometry analysis.

In preparation for flow cytometric analysis, the old media was removed from cell cultures and the cells were washed with PBS and detached by trypsination at 37ºC. Detached cells were collected in 2% FBS/PBS with or without 0.2% trypan blue. Samples were analysed by flow cytometry using a BD Biosciences FACSDiva and associated software. Untreated cells were defined as having 0% uptake.

References.

1. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O., *Science* **1991**, *254*, 1497-1500.

2. Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E., *Nature* **1993**, *365*, 566-568.

3. Kurreck, J., *European Journal of Biochemistry* **2003**, *270*, 1628-1644.

4. Larsen, H. J.; Bentin, T.; Nielsen, P. E., *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1999**, *1489*, 159-166.

5. Braasch, D. A.; Corey, D. R., *Biochemistry* **2002**, *41*, 4503-4510.

6. Cerqueira, L.; Azevedo, N. F.; Almeida, C.; Jardim, T.; Keevil, C. W.; Vieira, M. J., *International Journal of Molecular Sciences* **2008**, *9*, 1944-1960.

7. Malic, S.; Hill, K. E.; Hayes, A.; Percival, S. L.; Thomas, D. W.; Williams, D. W., *Microbiology* **2009**, *155*, 2603-11.

8. Chandler, D. P.; Stults, J. R.; Anderson, K. K.; Cebula, S.; Schuck, B. L.; Brockman, F. J., *Analytical Biochemistry* **2000**, *283*, 241-249.

9. Chandler, D. P.; Jarrell, A. E., *Analytical Biochemistry* **2003**, *312*, 182-190.

10. Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O., *Journal of Organic Chemistry* **1994**, *59*, 5767-5773.

11. Thomson, S. A.; Josey, J. A.; Cadilla, R.; Gaul, M. D.; Hassman, C. F.; Luzzio, M. J.; Pipe, A. J.; Reed, K. L.; Ricca, D. J.; Wiethe, R. W.; Noble, S. A., *Tetrahedron* **1995**, *51*, 6179-6194.

12. Breipohl, G.; Knolle, J.; Langner, D.; Omalley, G.; Uhlmann, E., *Bioorganic & Medicinal Chemistry Letters* **1996**, *6*, 665-670.

13. Stafforst, T.; Diederichsen, U., *European Journal of Organic Chemistry* **2007**, 681-688.

14. Diaz-Mochon, J. J.; Bialy, L.; Keinicke, L.; Bradley, M., *Chemical Communications* **2005**, 1384-1386.

15. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M., *Chemical Communications* **2006**, 3984-3986.

16. Pouchain, D.; Diaz-Mochon, J. J.; Bialy, L.; Bradley, M., *ACS Chemical Biology* **2007**, *2*, 810-818.

17. Bialy, L.; Diaz-Mochon, J. J.; Specker, E.; Keinicke, L.; Bradley, M., *Tetrahedron* **2005**, *61*, 8295-8305.

18. Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G., *International Journal of Peptide and Protein Research* **1991**, *37*, 487-493.

19. Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A., *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 10700-10704.

20. Brenner, S.; Lerner, R. A., *Proceedings of the National Academy of Sciences of the United States of America* **1992**, *89*, 5381-5383.

21. Still, W. C., *Accounts of Chemical Research* **1996**, *29*, 155-163.

22. Battersby, B. J.; Bryant, D.; Meutermans, W.; Matthews, D.; Smythe, M. L.; Trau, M., *Journal of the American Chemical Society* **2000**, *122*, 2138-2139.

23. Diaz-Mochon, J. J.; Bialy, L.; Bradley, M., *Organic Letters* **2004**, *6*, 1127-1129.

24. a) Winssinger, N.; Ficarro, S.; Schultz, P. G.; Harris, J. L., *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99, 11139-

11144. b) Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G., *Angewandte Chemie-International Edition* **2001**, 4, 3152-3155.

25. Harris, J. L.; Winssinger, N., *Chemistry-a European Journal* **2005**, *11*, 6792-6801.

26. Brenner, S.; Williams, S. R.; Vermaas, E. H.; Storck, T.; Moon, K.; McCollum, C.; Mao, J. I.; Luo, S. J.; Kirchner, J. J.; Eletr, S.; DuBridge, R. B.; Burcham, T.; Albrecht, G., *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 1665-1670.

27 Chhabra, S. R.; Hothi, B.; Evans, D. J.; White, P. D.; Bycroft, B. W.; Chan, W. C., *Tetrahedron Letters* **1998**, *39*, 1603-1606.

28. Gomez-Martinez, P.; Dessolin, M.; Guibe, F.; Albericio, F., *Journal of the Chemical Society-Perkin Transactions 1* **1999**, 2871-2874.

29. Jenner, J. R.; Buttle, D. J.; Dixon, A. K., *Annals of the Rheumatic Diseases* **1986**, *45*, 441-449.

30. Simmons, J. W.; Nordby, E. J.; Hadjipavlou, A. G., *European Spine Journal* **2001**, *10*, 192-202.

31. Huet, J.; Looze, Y.; Bartik, K.; Raussens, V.; Wintjens, R.; Boussard, P., *Biochemical and Biophysical Research Communications* **2006**, *341*, 620-626.

32. Bernstein, D. I.; Gallagher, J. S.; Grad, M.; Bernstein, I. L., *Journal of Allergy and Clinical Immunology* **1984**, *74*, 258-260.

33. Diaz-Mochon, J. J.; Planonth, S.; Bradley, M., *Analytical Biochemistry* **2009**, *384*, 101-105.

34. Ebata, M.; Yasunobu, K. T., *Journal of Biological Chemistry* **1962**, *237*, 1086-1094.

35. Baines, B. S.; Brocklehurst, K.; Carey, P. R.; Jarvis, M.; Salih, E.; Storer, A. C., *Biochemical Journal* **1986**, *233*, 119-129.

36. Fujii, N., *Origins of Life and Evolution of the Biosphere* **2002**, *32*, 103-127.

37. Bugg, T. D. H.; Walsh, C. T., *Natural Product Reports* **1992**, *9*, 199-215.

38. Asano, Y.; Lubbehusen, T. L., Journal of Bioscience and Bioengineering **2000**, 89, 295-306.

39. Helfman, P. M.; Bada, J. L., Proceedings of the National Academy of Sciences of the United States of America **1975**, 72, 2891-2894.

40. Helfman, P. M.; Bada, J. L., *Nature* **1976**, *262*, 279-281.

41. Fisher, G. H.; Garcia, N. M.; Payan, I. L.; Cadillaperezrios, R.; Sheremata, W. A.; Man, E. H., *Biochemical and Biophysical Research Communications* **1986**, *135*, 683-687.

42. Shapira, R.; Chou, C. H. J., *Biochemical and Biophysical Research Communications* **1987**, *146*, 1342-1349.

43. Roher, A. E.; Lowenson, J. D.; Clarke, S.; Wolkow, C.; Wang, R.; Cotter, R. J.; Reardon, I. M.; Zurcherneely, H. A.; Heinrikson, R. L.; Ball, M. J.; Greenberg, B. D., *Journal of Biological Chemistry* **1993**, *268*, 3072-3083.

44. Masters, P. M.; Bada, J. L.; Zigler, J. S., *Nature* **1977**, *268*, 71-73.

45. Fujii, N.; Muraoka, S.; Harada, K., *Biochimica Et Biophysica Acta* **1989**, *999*, 239-242.

46. Fujii, N.; Momose, Y.; Ishibashi, Y.; Uemura, T.; Takita, M.; Takehana, M., *Experimental Eye Research* **1997**, *65*, 99-104.

47. Hashimoto, A.; Oka, T., *Progress in Neurobiology* **1997**, *52*, 325-353.

48. Jilek, A.; Kreil, G., *Monatshefte Fur Chemie* **2008**, *139*, 1-5.

49. Csordas, A.; Michl, H., *Monatshefte Fur Chemie* **1970**, *101*, 182-189.

50. Simmaco, M.; Kreil, G.; Barra, D., *Biochimica Et Biophysica Acta-Biomembranes* **2009**, *1788*, 1551-1555.

51. Mignogna, G.; Simmaco, M.; Kreil, G.; Barra, D., *Embo Journal* **1993**, *12*, 4829-4832.

52. Faraci, W. S.; Walsh, C. T., *Biochemistry* **1988**, *27*, 3267-3276.

53. Walsh, C. T., *Journal of Biological Chemistry* **1989**, *264*, 2393-2396.

54. Wolosker, H.; Sheth, K. N.; Takahashi, M.; Mothet, J. P.; Brady, R. O.; Ferris, C. D.; Snyder, S. H., *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 721-725.

55. Fujii, N.; Harada, K.; Momose, Y.; Ishii, N.; Akaboshi, M., *Biochemical and Biophysical Research Communications* **1999**, *263*, 322-326.

56. Asano, Y.; Nakazawa, A.; Kato, Y.; Kondo, K., *Journal of Biological Chemistry* **1989**, *264*, 14233-14239.

57. Ozaki, A.; Kawasaki, H.; Yagasaki, M.; Hashimoto, Y., *Bioscience Biotechnology and Biochemistry* **1992**, *56*, 1980-1984.

58. Smal, M. A.; Dong, Z.; Cheung, H. T. A.; Asano, Y.; Escoffier, L.; Costello, M.; Tattersall, M. H. N., *Biochemical Pharmacology* **1995**, *49*, 567-574.

59. Kolb, H. C.; Finn, M. G.; Sharpless, K. B., *Angewandte Chemie-International Edition* **2001**, *40*, 2004-2021.

60. Codelli, J. A.; Baskin, J. M.; Agard, N. J.; Berozzi, C. R., *Journal of the American Chemical Society* **2008**, *130* , 11486-11493.

61. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., *Angewandte Chemie-International Edition* **2002**, *41*, 2596-2599.

62. Zhang, L.; Chen, X. G.; Xue, P.; Sun, H. H. Y.; Williams, I. D.; Sharpless, K. B.; Fokin, V. V.; Jia, G. C., *Journal of the American Chemical Society* **2005**, *127*, 15998-15999.

63. Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., *Journal of the American Chemical Society* **2004**, *126*, 15046-15047.

64. van Berkel, S. S.; Dirks, A. T. J.; Debets, M. F.; van Delft, F. L.; Cornelissen, J.; Nolte, R. J. M.; Rutjes, F., *Chembiochem* **2007**, *8*, 1504-1508.

65. Pieters, R. J.; Rijkers, D. T. S.; Liskamp, R. M. J., *QSAR & Combinatorial Science* **2007**, *26*, 1181-1190.

66. von Maltzahn, G.; Ren, Y.; Park, J. H.; Min, D. H.; Kotamraju, V. R.; Jayakumar, J.; Fogal, V.; Sailor, M. J.; Ruoslahti, E.; Bhatia, S. N., *Bioconjugate Chemistry* **2008**, *19*, 1570-1578.

67. Morales-Sanfrutos, J.; Ortega-Munoz, M.; Lopez-Jaramillo, J.; Hernandez-Mateo, F.; Santoyo-Gonzalez, F., *Journal of Organic Chemistry* **2008**, *73*, 7768-7771.

68. Morales-Sanfrutos, J.; Ortega-Munoz, M.; Lopez-Jaramillo, J.; Hernandez-Mateo, F.; Santoyo-Gonzalez, F., *Journal of Organic Chemistry* **2008**, *73*, 7772-7774.

69. Cassidy, M. P.; Raushel, J.; Fokin, V. V., *Angewandte Chemie-International Edition* **2006**, *45* (19), 3154-3157.

70. Govindaraju, T.; Jonkheijm, P.; Gogolin, L.; Schroeder, H.; Becker, C. F. W.; Niemeyer, C. M.; Waldmann, H., *Chemical Communications* **2008**, (32), 3723-3725.

71. Salic, A.; Mitchison, T. J., *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 2415-2420.

72. Herce, H. D.; Garcia, A. E., *Journal of Biological Physics* **2007**, *33,* 345-356.

73. Vives, E.; Brodin, P.; Lebleu, B., *Journal of Biological Chemistry* **1997**, *272,* 16010-16017.

74. Amand, H. L.; Fant, K.; Norden, B.; Esbjorner, E. K., *Biochemical and Biophysical Research Communications* **2008**, *371,* 621-625.

75. Joliot, A.; Pernelle, C.; Deagostinibazin, H.; Prochiantz, A., *Proceedings of the National Academy of Sciences of the United States of America* **1991**, *88,* 1864-1868.

76. Morris, M. C.; Vidal, P.; Chaloin, L.; Heitz, F.; Divita, G., *Nucleic Acids Research* **1997**, *25,* 2730-2736.

77. Morris, M. C.; Deshayes, S.; Heitz, F.; Divita, G., *Biology of the Cell* **2008**, *100,* 201-217.

78. Futaki, S.; Suzuki, T.; Ohashi, W.; Yagami, T.; Tanaka, S.; Ueda, K.; Sugiura, Y., *Journal of Biological Chemistry* **2001**, *276,* 5836-5840.

79. Vives, E.; Brodin, P.; Lebleu, B., *Journal of Biological Chemistry* **1997**, *272*, 16010-16017.

80. Derossi, D.; Joliot, A. H.; Chassaing, G.; Prochiantz, A., *Journal of Biological Chemistry* **1994**, *269,* 10444-10450.

81. Langel, U.; Pooga, M.; Kairane, C.; Zilmer, M.; Bartfai, T., *Regulatory Peptides* **1996**, *62,* 47-52.

82. Pooga, M.; Hallbrink, M.; Zorko, M.; Langel, U., *Faseb Journal* **1998**, *12,* 67-77.

83. Stewart, K. M.; Horton, K. L.; Kelley, S. O., *Organic & Biomolecular Chemistry* **2008**, *6,* 2242-2255.

84. Derossi, D.; Calvet, S.; Trembleau, A.; Brunissen, A.; Chassaing, G.; Prochiantz, A., *Journal of Biological Chemistry* **1996**, *271*, 18188-18193.

85. Duclohier, H., *European Biophysics Journal with Biophysics Letters* **2006**, *35*, 401-409.

86. Rothbard, J. B.; Garlington, S.; Lin, Q.; Kirschberg, T.; Kreider, E.; McGrane, P. L.; Wender, P. A.; Khavari, P. A., *Nature Medicine* **2000**, *6*, 1253-1257.

87. Jarvert, P.; Langel, K.; El-Andaloussi, S.; Langel, U., *Biochemical Society Transactions* **2007**, *35*, 770-774.

88. Turner, J. J.; Ivanova, G. D.; Verbeure, B.; Williams, D.; Arzumanov, A. A.; Abes, S.; Lebleu, B.; Gait, M. J., *Nucleic Acids Research* **2005**, *33*, 6837-6849.

89. Jiang, T.; Olson, E. S.; Nguyen, Q. T.; Roy, M.; Jennings, P. A.; Tsien, R. Y., *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 17867-17872.

90. Gros, E.; Deshayes, S.; Morris, M. C.; Aldrian-Herrada, G.; Depollier, J.; Heitz, F.; Divita, G., *Biochimica Et Biophysica Acta-Biomembranes* **2006**, *1758*, 384-393.

91. Taneja, N.; Davis, M.; Choy, J. S.; Beckett, M. A.; Singh, R.; Kron, S. J.; Weichselbaum, R. R., *Journal of Biological Chemistry* **2004**, *279*, 2273-2280.

92. Morris, M. C.; Depollier, J.; Mery, J.; Heitz, F.; Divita, G., *Nature Biotechnology* **2001**, *19*, 1173-1176.

93. Li, X.; Hummer, A.; Han, J.; Xie, M.; Melnik-Martinez, K.; Moreno, R. L.; Buck, M.; Mark, M. D.; Herlitze, S., *Journal of Biological Chemistry* **2005**, *280*, 23945-23959.

94. Buster, D.; McNally, K.; McNally, F. J., *Journal of Cell Science* **2002**, *115*, 1083-1092.

95. Svensen, N.; Diaz-Mochon, J. J.; Dhaliwal, K.; Planonth, S.; Dewar, M.; Armstrong, D.; Bradley, M., *Angewandte Chemie-International Edition* **2011**, *50*, 6133-6136.

96. Peretto, I.; Sanchez-Martin, R. M.; Wang, X. H.; Ellard, J.; Mittoo, S.; Bradley, M., *Chemical Communications* **2003**, 2312-2313.

97. Kaiser, E.; Colescot, R.L.; Bossinge, C.D.; Cook, P. I., *Analytical Biochemistry* **1970**, *34*, 595-598.

98. Fara, M. A.; Diaz-Mochon, J. J.; Bradley, M., *Tetrahedron Letters* **2006**, *47*, 1011-1014.

99. Chhabra, S. R.; Parekh, H.; Khan, A. N.; Bycroft, B. W.; Kellam, B., *Tetrahedron Letters* **2001**, *42*, 2189-2192.

100. Fischer, R.; Mader, O.; Jung, G.; Brock, R., *Bioconjugate Chemistry* **2003**, *14*, 653-660.

101. Svensen, N., *Unplublished PhD's thesis* **2011**, *University of Edinburgh,* 51-71.